

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:38:25 ; Search time 30.38 Seconds
(without alignments)
1104.157 Million cell updates/sec

Title: US-09-652-743A-12
Perfect score: 1559
Sequence: 1 MKVKKLIPSLVFGALVSP.....NGGPKATAAFVEINTAKAQ 302

ring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5
747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT *
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT *
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT *
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT *
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT *
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT *
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT *
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT *
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT *
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT *
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT *
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT *
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT *
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT *
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT *
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT *
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT *
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT *
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT *
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT *
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT *
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1559	100.0	302	22	AAE00016	Bacillus pumilus J
2	1506	96.6	303	22	AAE00014	Bacillus pumilus B
3	1118	71.7	215	22	AAE00196	Bacillus pumilus R
4	1117	71.6	215	22	AAE00197	Bacillus pumilus R
5	1117	71.6	215	22	AAE00200	Bacillus pumilus R
6	1116	71.6	215	22	AAE00199	Bacillus pumilus R
7	1110	71.2	215	22	AAE00198	Bacillus pumilus R
8	435.5	27.9	318	22	AAE00015	Bacillus lichenifo
9	420.5	27.0	316	13	AA23730	Protease Blase. S
10	419.5	26.9	316	22	AAE00011	Bacillus lichenifo
11	417	26.7	314	22	AAE00013	Bacillus lichenifo

12	412.5	26.5	222	22	AAE00075	Bacillus lichenifo
13	411.5	26.4	222	22	AAE00036	Bacillus lichenifo
14	411.5	26.4	222	22	AAE00116	Bacillus lichenifo
15	411.5	26.4	222	22	AAE00118	Bacillus lichenifo
16	410.5	26.3	222	22	AAE00071	Bacillus lichenifo
17	409.5	26.3	222	22	AAE00046	Bacillus lichenifo
18	409.5	26.3	222	22	AAE00048	Bacillus lichenifo
19	409	26.2	221	22	AAE00187	Bacillus lichenifo
20	409	26.2	222	22	AAE00053	Bacillus lichenifo
21	409	26.2	222	22	AAE00190	Bacillus lichenifo
22	408.5	26.2	222	22	AAE00092	Bacillus lichenifo
23	408.5	26.2	222	22	AAE00137	Bacillus lichenifo
24	407.5	26.1	222	22	AAE00047	Bacillus lichenifo
25	407.5	26.1	222	22	AAE00049	Bacillus lichenifo
26	407.5	26.1	222	22	AAE00140	Bacillus lichenifo
27	407	26.1	222	22	AAE00193	Bacillus lichenifo
28	406.5	26.1	222	22	AAE00033	Bacillus lichenifo
29	406.5	26.1	222	22	AAE00076	Bacillus lichenifo
30	406.5	26.1	222	22	AAE00107	Bacillus lichenifo
31	406.5	26.1	222	22	AAE00182	Bacillus lichenifo
32	406.5	26.1	222	22	AAE00205	Bacillus lichenifo
33	406	26.0	222	22	AAE00112	Bacillus lichenifo
34	405.5	26.0	222	12	AA14159	Serine protease SP
35	405.5	26.0	222	13	AA26467	Sequence of a seri
36	405.5	26.0	222	22	AAE00050	Bacillus lichenifo
37	405.5	26.0	222	22	AAE00052	Bacillus lichenifo
38	405.5	26.0	222	22	AAE00073	Bacillus lichenifo
39	405.5	26.0	222	22	AAE00077	Bacillus lichenifo
40	405.5	26.0	222	22	AAE00078	Bacillus lichenifo
41	405.5	26.0	222	22	AAE00108	Bacillus lichenifo
42	405.5	26.0	222	22	AAE00109	Bacillus lichenifo
43	405.5	26.0	222	22	AAE00170	Bacillus lichenifo
44	405.5	26.0	222	22	AAE00172	Bacillus lichenifo
45	405.5	26.0	222	22	AAE00184	Bacillus lichenifo

ALIGNMENTS

RESULT 1
AAE00016
ID AAE00016 standard; Protein; 302 AA.

XX AC AAE00016;

XX DT 31-MAY-2001 (first entry)

XX DE Bacillus pumilus JA96 RP-II protease.

XX XX Residual protease II; RP-II; additive; cleaning composition; detergent.

XX OS Bacillus pumilus JA96.

PH Key Location/Qualifiers

FT Peptide 1..25 /label= Signal_peptide

FT Peptide 26..87 /label= Pro_peptide

FT FT /note= "Pro sequence of RP-II protease"

FT Protein 88..302 /note= "Bacillus pumilus JA96 mature RP-II protease"

XX WO200116285-A2.
XX 08-MAR-2001.

XX 31-AUG-2000; 2000WO-DK00476.

XX 31-AUG-1999; 99DK-0001212

XX 20-OCT-1999; 99DK-0001500.

XX (NOVO) NOVOZYMES AS.

PI Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
PI Flensted Lassen S;
XX WPI; 2001-226680/23.
DR N-PSDB; AAD03004.
XX
XX Novel RP-II type protease and its variants useful as constituents in
PT detergent compositions, additives and cleaning compositions -
XX
XX Claim 1; Page 119-120; 132pp; English.
XX
XX The present sequence is Bacillus pumilus JA96 RP-II (Residual
CC protease II). RP-II protease is useful as a constituent in additives,
CC detergent compositions and other cleaning compositions, optionally in
CC combination with other enzymes such as proteases, lipases, cellulases,
CC amylases, peroxidases or oxidases. The variants of RP-II have improved
CC properties such as substrate specificities, catalytic rate, stability,
CC especially towards the action of proteolytic enzymes and improved
CC resistance towards peroxidase.
XX
XX Sequence 302 AA;
XX
XX Query Match 100.0%; Score 1559; DB 22; Length 302;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-116; Mismatches 0; Gaps 0;
XX Matches 302; Conservative 0; Indels 0; Gaps 0;
XX
XX 1 MKKKVKKLIPSLVFGALSVPSFAHAASDVLTSYDMVTSYGKVISSADFNHDMKTPSSF 60
XX Db 1 mkkvkklipslvfgalsvpsfahaaadsvltsydmvtsydgkviissadfnhdmktpsf 60
XX
XX 61 DKVDLSTTIGEKVKPLTTLTKDFQTVIGDGRKVTNTRVAPYNSIAYITFGSSCT 120
XX Db 61 dkvdlssttiguekvkplttlytkdfqtkvvgdgrtkvtntrvapynsiayitfgssct 120
XX
XX 121 GTLIAPNKLITNGHCYNTATRSYSKAGSVYPCGNDSTAVNGSANTFEYVPSGYINTGA 180
XX Db 121 gtlipnklitngchcynatrsyskagsvypcgndstavnsgsanmtfeyvpsgyintga 180
XX
XX 181 SQYDFAVIKTDNIGNTVGRSIRQVTLNLTGTTIKISGYPGDKMRSTGKYSQWEMSGPVT 240
XX Db 181 sqydfaviktdnigntvgrsirqvtnlntgttikisgypgdkmrstgkysqwemsgpvt 240
XX
XX 241 REDTNLAYTIDTFSGNSGAMLDQNOQIVGVHNAGYSNGTINGGPKATAAFVEFINYAK 300
XX Db 241 redtnlaytiddtfsgnsgamldqnoqivgvhnagysngtinggpkataafvefinyak 300
XX
XX 301 AQ 302
XX Db 301 aq 302
XX
XX .T 2
XX 0014
XX AAE00014 standard; Protein; 303 AA.
XX
XX AAE00014;
XX AC AAE00014;
XX
XX 31-MAY-2001 (first entry)
XX
XX Bacillus pumilus B032 RP-II protease.
XX
XX Residual protease II; RP-II; additive; cleaning composition; detergent.
XX
XX Bacillus pumilus B032.
XX
XX Key Location/Qualifiers
XX Peptide 1..26
XX FT /label= Signal_peptide
XX FT 27..88
XX FT /label= Pro_peptide
XX FT /note= "Pro sequence of RP-II protease"
XX FT 89..303

FT WO200116285-A2.
XX 08-MAR-2001.
XX
XX 31-AUG-2000; 2000WO-DK00476.
XX
XX 31-AUG-1999; 99DK-0001212 ✓
XX 20-OCT-1999; 99DK-0001500 ✓
XX
XX (NOVO) NOVOZYMES AS.
XX
XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
XX Flensted Lassen S;
XX
XX WPI; 2001-226680/23.
XX DR N-PSDB; AAD03002.
XX
XX Novel RP-II type protease and its variants useful as constituents in
PT detergent compositions, additives and cleaning compositions -
XX
XX Claim 1; Page 112-113; 132pp; English.
XX
XX The present sequence is Bacillus pumilus B032 RP-II (Residual
CC protease II). RP-II protease is useful as a constituent in additives,
CC detergent compositions and other cleaning compositions, optionally in
CC combination with other enzymes such as proteases, lipases, cellulases,
CC amylases, peroxidases or oxidases. The variants of RP-II have improved
CC properties such as substrate specificities, catalytic rate, stability,
CC especially towards the action of proteolytic enzymes and improved
CC resistance towards peroxidase.
XX
XX Sequence 303 AA;
XX
XX Query Match 96.6%; Score 1506; DB 22; Length 303;
XX Best Local Similarity 95.7%; Pred. No. 2.6e-112; Mismatches 7; Indels 0; Gaps 0;
XX Matches 289; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
XX
XX 1 MKKKVKKLIPSLVFGALSVPSFAHAASDVLTSYDMVTSYGKVISSADFNHDMKTPSSF 60
XX Db 2 mkkvkklipslvfgalsvpsfahaaadsvltsydmvtsydgkviissadfnhdmktpsf 61
XX
XX 61 DKVDLSTTIGEKVKPLTTLTKDFQTVIGDGRKVTNTRVAPYNSIAYITFGSSCT 120
XX Db 62 dkvdlssttiguekvkplskylkdfqtkvvgdgrtkvtntrvapynsiayitfgssct 121
XX
XX 121 GTLIAPNKLITNGHCYNTATRSYSKAGSVYPCGNDSTAVNGSANTFEYVPSGYINTGA 180
XX Db 122 gtlipnklitngchcynatrsyskagsvypcgndstavnsgsanmtfeyvpsgyintga 181
XX
XX 181 SQYDFAVIKTDNIGNTVGRSIRQVTLNLTGTTIKISGYPGDKMRSTGKYSQWEMSGPVT 240
XX Db 182 sqydfaviktdnigntvgrsirqvtnlntgttikisgypgdkmrstgkysqwemsgpvt 241
XX
XX 241 REDTNLAYTIDTFSGNSGAMLDQNOQIVGVHNAGYSNGTINGGPKATAAFVEFINYAK 300
XX Db 242 redtnlaytiddtfsgnsgamldqnoqivgvhnagysngtinggpkataafvefinyak 301
XX
XX 301 AQ 302
XX Db 302 aq 303
XX
XX RESULT 3
XX AAE00196
XX ID AAE00196 standard; Protein; 215 AA.
XX AC AAE00196;
XX
XX 31-MAY-2001 (first entry)
XX
XX

116 097X
A/88

116 097X
A/88

QY 268 QIVGVHNAGSYNGTINGGPKATAAFVEFINYAKAQ 302
 Db 181 qivgvhnagsyngtinggpkataafvefinyakaq 215

RESULT 4
 ID AAE00197 standard; Protein; 215 AA.
 XX AC AC
 XX AC AC
 XX DT DT
 XX XX
 DE Bacillus pumilus RP-II protease variant (E155G).
 XX
 KW Residual protease II; RP-II; additive; cleaning composition; detergent;
 KW mutant; mutain; variant.
 OS Bacillus pumilus JA96.
 OS Synthetic.
 XX
 FF Key Location/Qualifiers
 FT Misc-difference 155
 FT FT /note= "Wild type Glu substituted with Gly"
 XX
 PN WO200116285-A2.
 XX
 XX 08-MAR-2001.
 XX
 XX 31-AUG-2000; 2000WO-DK00476.
 XX
 XX 31-AUG-1999; 99DK-00012124✓
 XX 20-OCT-1999; 99DK-0001500.
 XX
 XX (NOVO) NOVOZYMES AS.
 XX
 XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
 PI Flensted Lassen S;
 XX
 XX WPI: 2001-226680/23.
 XX
 XX Novel RP-II type protease and its variants useful as constituents in
 PT detergent compositions, additives and cleaning compositions -
 PT
 XX Example 3; Page -: 132pp; English.
 XX
 XX The present sequence is Bacillus pumilus (JA96) RP-II protease
 CC variant (E155G).
 CC The patent discloses RP-II (Residual protease II) protein, their
 CC corresponding nucleic acid sequences and variants. RP-II protease
 CC is useful as a constituent in additives, detergent compositions and
 CC other cleaning compositions, optionally in combination with other
 CC enzymes such as proteases, lipases, cellulases, amylases, peroxidases
 CC or oxidases. The variants of RP-II have improved properties such as
 CC substrate specificities, catalytic rate, stability, especially towards
 CC the action of proteolytic enzymes and improved resistance towards
 CC peroxidase.

peroxide. This sequence is not shown in the specification but is derived from the mature B. licheniformis RP11 protein (SEQ ID NO: 2) shown in page 101-102 of the specification (AAE00011). The specification also refers to following variants: (a) V1F, (b) D7G-TL25S+EL52G+NI182I, (c) SI19A-S144T+Q157R+SI175, (d) H141A, (e) E152A+Q122R, (f) NI82T+CI185A, (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a) and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in (d), Gln residue in (e) and (h) do not match with the B. licheniformis native wild type RP11 protease sequence.

Sequence 215 AA;

Query Match	71.6%	Score 1117;	DB 22;	Length 215;
Best Local Similarity	99.5%;	Pred. No. 1.8e-81;		
Matches 214;	Conservative	0;	Mismatches 1;	Indels 0; Gaps

CC variant (D156A).
 CC The patent discloses RP-II (Residual protease II) protein, their
 CC corresponding nucleic acid sequences and variants. RP-II protease
 CC is useful as a constituent in additives, detergent compositions and
 CC other cleaning compositions, optionally in combination with other
 CC enzymes such as proteases, lipases, cellulases, amylases, peroxidases
 CC or oxidases. The variants of RP-II have improved properties such as
 CC substrate specificities, catalytic rate, stability, especially towards
 CC the action of proteolytic enzymes and improved resistance towards
 CC peroxidase.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in
 CC page 101-102 of the specification (AAE00011). The specification also
 CC refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,
 CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,
 CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)
 CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in
 CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis
 CC native wild type RPII protease sequence.
 XX
 SQ Sequence 215 AA;

Query Match 71.6%; Score 1116; DB 22; Length 215;
 Best Local Similarity 99.5%; Pred. No. 2.2e-81;
 Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 147
 DB 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 60
 QY 148 GSVYPMNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 207
 DB 61 GSVYPMNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 120
 QY 208 NLGTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAYTYTIDTFSGNSGSMALDQNG 267
 DB 121 NLGTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAYTYTIDTFSGNSGSMALDQNG 180
 QY 268 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 302
 DB 181 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 215

RESULT 7
 AAE00198
 AAE00198 standard; Protein: 215 AA.
 AAE00198;
 31-MAY-2001 (first entry)
 Bacillus pumilus RP-II protease variant (E155A+D156A).
 Residual protease II: RP-II; additive; cleaning composition; detergent;
 mutant; mutein; variant.
 Bacillus pumilus JA96.
 Synthetic.

Key Location/Qualifiers
 FT Misc-difference 155
 FT /note= "Wild type Glu substituted with Ala"
 FT Misc-difference 156
 FT /note= "Wild type Asp substituted with Ala"
 XX

PN WO200116285-A2
 XX
 XX 08-MAR-2001.
 XX
 XX 31-AUG-2000; 2000WO-DK00476.
 XX
 XX 31-AUG-1999; 99DK-0001212. ✓

PR 20-OCT-1999; 99DK-0001500. ✓
 XX (NOVO) NOVOZYMES AS.
 XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
 PI Flensted Lassen S;
 XX WPI: 2001-226680/23.
 XX Novel RP-II type protease and its variants useful as constituents in
 PT detergent compositions, additives and cleaning compositions -
 PS Example 3; Page -: 132pp; English.
 XX The present sequence is Bacillus pumilus (JA96) RP-II protease
 CC variant (E155A+D156A).
 CC The patent discloses RP-II (Residual protease II) protein, their
 CC corresponding nucleic acid sequences and variants. RP-II protease
 CC is useful as a constituent in additives, detergent compositions and
 CC other cleaning compositions, optionally in combination with other
 CC enzymes such as proteases, lipases, cellulases, amylases, peroxidases
 CC or oxidases. The variants of RP-II have improved properties such as
 CC substrate specificities, catalytic rate, stability, especially towards
 CC the action of proteolytic enzymes and improved resistance towards
 CC peroxidase.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in
 CC page 101-102 of the specification (AAE00011). The specification also
 CC refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,
 CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,
 CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)
 CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in
 CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis
 CC native wild type RPII protease sequence.
 XX
 SQ Sequence 215 AA;

Query Match 71.2%; Score 1110; DB 22; Length 215;
 Best Local Similarity 99.1%; Pred. No. 6.7e-81;
 Matches 213; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 147
 DB 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 60
 QY 148 GSVYPMNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 207
 DB 61 GSVYPMNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 120
 QY 208 NLGTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAYTYTIDTFSGNSGSMALDQNG 267
 DB 121 NLGTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAYTYTIDTFSGNSGSMALDQNG 180
 QY 268 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 302
 DB 181 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 215

RESULT 8
 AAE00015
 ID AAE00015 standard; Protein: 318 AA.
 XX
 AC AAE00015;
 XX
 DT 31-MAY-2001 (first entry)
 XX
 XX Bacillus licheniformis CDJ31 RP-II protease.
 DE Residual protease II: RP-II; additive; cleaning composition; detergent.
 KW
 XX
 OS Bacillus licheniformis CDJ31.
 XX

Db 121 ssigsgctgmigpaktvatagchlydtssgafagtatsvprngtstpygsvkstrypis 180

QY 174 GYINTGASQYDFAVIKTDTNIGTNGVRSIROVT-NLTGTTIKISGYPGDKMRSTGKVSQ 232

Db 181 gw-rsgntnydgaieisepigtvgyfysyttsslvgtvtvisgpgdktagt---q 235

QY 233 WEMSGPVTRDTNLYATTIDTFSGNSGSAMLQNOQ-----IVGVHAG-YSNGTIN 283

Db 236 wqhsqplaisetyklyqamdyggsgspvfedqssrtncsgpcslavhtngvgygssyn 295

QY 284 GPKAT 289

Db 296 rgrit 301

TLT 10

0011

AAE00011 standard; Protein; 316 AA.

AAE00011;

31-MAY-2001 (first entry)

Bacillus licheniformis (BLC) RP-II protease.

Residual protease II; RP-II; additive; cleaning composition; detergent.

Bacillus licheniformis.

Key Location/Qualifiers

Peptide 1..31

Peptide /label= Signal_peptide

Peptide 32..94

Peptide /label= Pro_peptide

Protein /note= "Pro sequence of RP-II protease"

Protein 95..316

Protein /note= "Bacillus licheniformis mature RP-II protease"

WO200116285-A2.

08-MAR-2001.

31-AUG-2000; 2000WO-DK00476. ✓

31-AUG-1999; 99DK-0001212. ✓

20-OCT-1999; 99DK-0001500. ✓

(NOVO) NOVOZYMES AS.

Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB; Flensted Lassen S;

WPI: 2001-226680/23.

N-PSDB; AAD02999.

Novel RP-II type protease and its variants useful as constituents in detergent compositions, additives and cleaning compositions

Claim 1; Page 101-102; 132pp; English.

The present sequence is Bacillus licheniformis (BLC) RP-II (Residual protease II). RP-II protease is useful as a constituent in additives, detergent compositions and other cleaning compositions, optionally in combination with other enzymes such as proteases, lipases, cellulases, amylases, peroxidases or oxidases. The variants of RP-II have improved properties such as substrate specificities, catalytic rate, stability, especially towards the action of proteolytic enzymes and improved resistance towards peroxidase.

Sequence 316 AA;

Query Match 26.9%; Score 419.5; DB 22; Length 316;

Best Local Similarity 33.9%; Pred. No. 1.5e-25;

Matches 106; Conservative 49; Mismatches 117; Indels 41; Gaps 11;

QY 2 KVKK-LIPSLVFGALSV---PFAHAAASDVLTS-----YD-MVTSQKVIS 46

Db 5 ksvkrlitgligisiyslgmhpaqaapsbhtpvsdpsyaetsvtdphiksdqgily 64

QY 47 SADFHNDKTPSFDDKVDLSSTIGEKVKPLTTLTKDFQTKVVIIGDDGRTKVTNTRVAPY 106

Db 65 skaftgtgkvetkek-----aekspaka---pysiksvigsdrrtrvnttapy 113

QY 107 NSIAYITFGSSCTCTLIAPNKILTNCHCYVNTATRSYSAKGSVYPCGMNDSTAVNGSANM 166

Db 114 raivhissigsgctgmigpaktvatagchlydtssgafagtatsvprngtstpygsvk 173

QY 167 TEFYVPSSGYINTGASQYDFAVIKTDTNIGTNGVRSIROVT-NLTGTTIKISGYPGDKMR 225

Db 174 tryfipsgw-rsgntnydgaieisepigtvgyfysyttsslvgtvtvisgpgdkt 232

QY 226 STCKVQWEMSGPVTRDTNLYATTIDTFSGNSGSAMLQNOQ-----IVGVHAG- 276

Db 233 gt----qwhsgplaisetyklyqamdyggsgspvfedqssrtncsgpcslavhtngv 288

QY 277 YSNGTINGGPKAT 289

Db 289 yggssynrgtrit 301

RESULT 11

AAE00013

ID AAE00013 standard; Protein; 314 AA.

XX AC AAE00013;

XX DT 31-MAY-2001 (first entry)

XX DE Bacillus licheniformis AC116 RP-II protease.

XX KW Residual protease II; RP-II; additive; cleaning composition; detergent.

XX OS Bacillus licheniformis.

XX FH Key Location/Qualifiers

FT Peptide 1..29

FT Peptide /label= Signal_peptide

FT Peptide 30..92

FT Peptide /label= Pro_peptide

FT Protein /note= "Pro sequence of RP-II protease"

FT Protein 93..314

FT Protein /note= "Bacillus licheniformis AC116 mature RP-II protease"

XX WO200116285-A2.

XX 08-MAR-2001.

XX 31-AUG-2000; 2000WO-DK00476.

XX 31-AUG-1999; 99DK-0001212. ✓

XX 20-OCT-1999; 99DK-0001500. ✓

XX (NOVO) NOVOZYMES AS.

XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB; Flensted Lassen S;

XX WPI: 2001-226680/23.

XX N-PSDB; AAD03001.

XX Novel RP-II type protease and its variants useful as constituents in detergent compositions, additives and cleaning compositions

XX Claim 1; Page 109-110; 132pp; English.

[illegible]

Db 121 slvgttvtisgpgdktagt----qwhsgpialsetykiqyamdyggsgspvfeqs 176
Qy 268 Q-----IVGVHAG-YSNGTINGGPKAT 289
Db 177 srtnsgpcslavhtngvygssynrgtrlt 207

RESULT 15 -
AAE00118
ID AAE00118 standard; Protein; 222 AA.
XX AC AAE00118;
XX DT 31-MAY-2001 (first entry)
XX DE Bacillus licheniformis RP-II protease variant (T128K).
XX FW Residual protease II; RP-II; additive; cleaning composition; detergent;
mutant; mutein; variant.
Bacillus licheniformis.
Synthetic.

Key Location/Qualifiers
Misc-difference 128
/note= "Wild type Thr substituted with Lys"
W0200116285-A2.
08-MAR-2001.
31-AUG-2000; 2000WO-DK00476.
31-AUG-1999; 99DK-0001212.6 ✓
20-OCT-1999; 99DK-0001500.
(NOVO) NOVOZYMES AS.
Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
Flensted Lassen S;
WPI; 2001-226680/23.
Novel RP-II type protease and its variants useful as constituents in
detergent compositions, additives and cleaning compositions
Example 3; Page -: 132pp; English.
The present sequence is Bacillus licheniformis (BLC) RP-II protease
variant (T128K).
The patent discloses RP-II (Residual protease II) protein, their
corresponding nucleic acid sequences and variants. RP-II protease
is useful as a constituent in additives, detergent compositions and
other cleaning compositions, optionally in combination with other
enzymes such as proteases, lipases, cellulases, amylases, peroxidases
or oxidases. The variants of RP-II have improved properties such as
substrate specificities, catalytic rate, stability, especially towards
the action of proteolytic enzymes and improved resistance towards
peroxidase.
Note: This sequence is not shown in the specification but is derived
from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in
page 101-102 of the specification (AAE00011). The specification also
refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,
(c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,
(g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)
and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in
(d), Gln residue in (e) and (h) do not match with the B. licheniformis
native wild type RPII protease sequence.
XX Sequence 222 AA;
SQ

Query Match 26.4%; Score 411.5; DB 22; Length 222;
Best Local Similarity 39.8%; Pred. No. 4e-25;
Matches 84; Conservative 36; Mismatches 76; Indels 15; Gaps 5;
Qy 89 VIGDDGRTKVTNTRVAPYNSIAYITFGSSCTGLIAPNKILTNCHVCYNTATRSYSAKG 148
Db 2 vigsddtrvtunttaypyralvhisssigscgwmigpkkvtatagahciydissysfagta 61
Qy 149 SVYFGMNDSTAVNGSANNTEFYVPFGYINTGASQYDFAVIKTD7NIGNTVGYRSIROVT- 207
Db 62 tvspgrngtspypgsvkstryfipsgw-rsgntnydygaielsepignvtvgfygysyts 120
Qy 208 NLGCTTIKISGPGDKWRSTGKVSQWEMSGPVTREDTNLAYIYTTIDTFSGNSGSAMLQNO 267
Db 121 slvgttvkisgypgdktagt----qwhsgpialsetykiqyamdyggsgspvfeqs 176
Qy 268 Q-----IVGVHAG-YSNGTINGGPKAT 289
Db 177 srtnsgpcslavhtngvygssynrgtrlt 207

Search completed: June 21, 2002, 18:41:56
Job time: 211 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:42:01 ; Search time 30.24 Seconds
(without alignments)
1727.661 Million cell updates/sec

Title: US-09-652-743A-12

Perfect score: 1559

Sequence: 1 MKVKVKKLPSLLVFGALSVP.....NGGPKATAAFVEFINAKAQ 302

ing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

urched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1501	46.3	303	2 Q9EXR9	Q9exr9 bacillus in
2	341	21.9	358	16 Q931E7	Q931e7 staphylococ
3	254	16.3	364	16 Q98G17	Q98g17 rhizobium l
4	227.5	14.6	639	16 Q921B9	Q921b9 rhizobium m
5	214	13.7	316	2 Q9FBG1	Q9fbg1 staphylococ
6	202.5	13.0	188	16 Q99R42	Q99r42 staphylococ
7	192.5	12.3	357	2 Q04186	Q04186 staphylococ
8	189.5	12.2	342	16 Q99V45	Q99v45 staphylococ
9	189	12.1	217	2 Q9AJX0	Q9ajx0 staphylococ
10	187.5	12.0	271	16 Q987W6	Q987w6 enterococcu
11	179.5	11.5	284	2 Q47809	Q47809 enterococcu
12	174.5	11.2	240	2 Q9KH50	Q9kh50 staphylococ
13	173.5	11.1	240	16 Q53781	Q53781 staphylococ
14	161	10.3	238	2 Q9FD07	Q9fd07 staphylococ
15	156	10.0	239	2 Q9KH49	Q9kh49 staphylococ
16	156	10.0	330	16 Q9KLE3	Q9kle3 vibrio chol

097sh5 streptococ	153.5	9.8	285	16	097SH5
09kh47 staphylococ	153	9.8	238	16	09KH47
053782 staphylococ	150	9.6	239	16	053782
09fd08 staphylococ	144.5	9.3	235	2	09FD08
09kh08 staphylococ	144	9.2	268	2	09KH08
Q9vme6 drosophila	144	9.2	272	5	Q9VME6
Q9kh46 staphylococ	141.5	9.1	239	2	09KH46
099t61 staphylococ	141.5	9.1	239	16	099T61
099t60 staphylococ	139.5	8.9	235	16	099T60
09kh48 staphylococ	137.5	8.8	239	16	09KH48
Q00344 cochliobolu	134.5	8.6	261	3	Q00344
099r41 staphylococ	133.5	8.6	188	16	099R41
076602 caenorhabdi	130	8.3	1275	5	076602
018443 helicoverpa	129.5	8.3	276	5	018443
095084 homo sapien	129.5	8.3	383	4	095084
054137 saccharopol	128.5	8.2	272	2	054137
09kh51 staphylococ	127.5	8.2	235	2	09KH51
018445 helicoverpa	127.5	8.2	295	5	018445
018450 helicoverpa	126.5	8.1	295	5	018450
09d6x6 mus musculu	126.5	8.1	382	11	09D6X6
09y842 metarhizium	125	8.0	256	3	09Y842
09v519 drosophila	125	8.0	778	5	09V519
09vz58 drosophila	124.5	8.0	611	5	09VZ58
026996 methanother	124	8.0	698	17	026996
018655 plodia inte	123.5	7.9	282	5	018655
09ngv4 heliothis v	122.5	7.9	292	5	09NGY4
018444 helicoverpa	122.5	7.9	292	5	018444
09xy51 ctencephal	121.5	7.8	256	5	09XY51
Q9nh10 agrotis ips	121.5	7.8	287	5	Q9NH10

ALIGNMENTS

RESULT 1					
Q9EXR9	Q9EXR9	PRELIMINARY;	PRT;	303 AA.	
AC	Q9EXR9;				
DT	01-MAR-2001 (Tremblrel. 16, Created)				
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	GLUTAMYL-ENDOPEPTIDASE.				
OS	Bacillus intermedius				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID:1400;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-99170141; PubMed-10071925;				
RA	Rebrikov D.V., Akimkina T.V., Shevelev A.B., Demidyuk I.V.,				
RA	Bushneva A., Kostrov S.V., Chestukhina G.G., Stepanov V.M.;				
RT	"Bacillus intermedius glutamyl endopeptidase. Molecular cloning and				
RT	nucleotide sequence of the structural gene."				
RL	J. Protein Chem. 18:21-25(1999).				
DR	EMBL; Y15136; CAC17594.1;				
DR	InterPro; IPR000126; Ser_proteas_V8.				
DR	InterPro; IPR001254; Trypsin.				
DR	Pfam; PF00089; trypsin; 1				
DR	PRINTS; PRO0839; V8PROTEASE.				
DR	SMART; SM00020; Tryp_SPC; 1.				
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.				
KW	Hydrolase; Serine protease.				
SQ	SEQUENCE 303 AA; 32343 MW; F2E37B5B07781D59 CRC64;				

Query Match 96.3%; Score 1501; DB 2; Length 303;
Best Local Similarity 96.0%; Pred. No. 6e-91;
Matches 290; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy	1	MKKVKKLPSLLVFGALSVPFAHAASDVLTSYDWTSGDKVISSADFNDMKTPSPF 60	
Db	2	MKKVKKLPSLLVFGALSVPFAHAASDVLTSYDWTSGDKVISSADFNDMKTPSPF 61	

Native

QY 61 DKVDDLSSTIGEKVKPLTYLKDFQFVWIGDGRKTKVTRVAPYNSIAYITFGSSCT 120
Db 62 DKVDDLSSTIGEKVKPLTYLKDFQFVWIGDGRKTKVTRVAPYNSIAYITFGSSCT 121
QY 121 GTLIAPNKILTNCHGVNATRSYSAKSGVYPCGMDSTAVNGSANNTEYVPSGYINTCA 180
Db 122 GTLIAPNKILTNCHGVNATRSYSAKSGVYPCGMDSTAVNGSANNTEYVPSGYINTCA 181
QY 181 SQYDFAVIKTDNIGNTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPYT 240
Db 182 SQYDFAVIKTDNIGNTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPYT 241
QY 241 REDTNLAYITIDTFSGNSGSMALDQNOQIVGVHNGYNGTNGGPKATAAFVEFINYAK 300
Db 242 REDTNLAYITIDTFSGNSGSMALDQNOQIVGVHNGYNGTNGGPKATAAFVEFINYAK 301
QY 301 AQ 302
Db 302 AQ 303
T 2
QY 931E7 PRELIMINARY; PRT; 358 AA.
AC Q931E7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SAV2596.
GN SAV2596.
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003365; BAB58758.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 358 AA; 41086 MW; 5AB5B04202583B08 CRC64;

ry Match 21.9%; Score 341; DB 16; Length 358;
t Local Similarity 32.6%; Pred. No. 7.2e-15;
Matches 91; Conservative 51; Mismatches 111; Indels 26; Gaps 9;

QY 30 VLTSY---DMVTSQKVISSAFHNDKTPS-SFDKVDLSSTIGEKVKPLTYLKDFQ 85
Db 15 VIVSSYGESEFVLADNNPIVNSGSDNNFESGTTIERNQNELAEVGDYSTPL----- 66

QY 86 TKVVGDDGRKTKVTRVAPYNS--IAYITFGSSC---TGTLIAPNKILTNCHGVNTPA 140
Db 67 TERVFGKQDQRTVNNLTQRPYKGTGTVLLNNFTSNRRYKGTGTMGCKDIIVLTAHNVYSD 126

QY 141 TRSYSAKSGVYPCGMDSTAVNGSANNTEYVPSGYINTCASOYDFAVIKTDNIGNTVGY 200
Db 127 DKGWAKIDVYAGVNGQTTIGKAFSHKFFVSKTINNAPTREDTATIKLSNLGNKTYG 186

QY 201 RSIRQVNTLT-GTTIKISGYPGDKMRSTGKVSQWEMSGPVTRDNTNLAYITIDTFSGNSG 259
Db 187 LTLN--THLSKGENIEISGPGDKSDN----RQYKKGKLESFDENEMYTYVDTFSGQSG 240

QY 260 SAMLDQNOQIVGVHNGYNGTNGGPKATAAFVEFINY 298

Db 241 SAIRDSKNNIIGHVAYGRYNNH--NSGVRINDLKLDYINY 277
RESULT 3
Q98G17 PRELIMINARY; PRT; 364 AA.
AC Q98G17;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE GLUTAMYL ENDOPEPTIDASE (EC 3.4.21.19).
GN MLL3306.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003001; BAB50229.1; --
DR InterPro; IPR00126; Ser_proteas_v8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; Tryp_SPC; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 364 AA; 39034 MW; 13BE653270E7CDD3 CRC64;

Query Match 16.3%; Score 254; DB 16; Length 364;
Best Local Similarity 30.3%; Pred. No. 3.7e-09;
Matches 70; Conservative 38; Mismatches 95; Indels 28; Gaps 7;

QY 89 VIGDGRKTKVTRVAPYNSIAYI-----TFGSSCTGTLIAPNKILTNCHGVNTPATR 142
Db 136 VFGPDDREQVRNTKTYPFSAIGYLEAKSAKTSYGSCSATLIGPRTVLTAAHCLYSHEDK 195

QY 143 SYSAGSVYPCGMDSTAVN---GSANMTEFYVPSGYINTGASOY-----DFAVIKTDN 193
Db 196 DMLSEYLFVPLNGSTADDAPEGAFYTESAYVLOGFIDNYQGYGVSVPMDLGIITLAKD 255

QY 194 IGNTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRDNTNLAY--YTI 251
Db 256 VGTNLCWLGANYDDLGDFETANLVGYPGDKPMGT----HWKASCEVHAENIAPEYFQYDC 311

QY 252 DTFSGNSGSMALDQNOQIVGVHNGYNGTNGGPKATAAFVEFIN 297
Db 312 DTFPGSSGVVAYDTKSKRIITGVNVA--ESPDAVTLAVLNAAVQWIN 360

RESULT 4
Q92LB9 PRELIMINARY; PRT; 639 AA.
AC Q92LB9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE PROTEASE TRANSMEMBRANE PROTEIN.
GN SMC03769.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;

RT Mesorhizobium loti. #;
 DR DNA Res. 7:331-338(2000).
 DR EMBL: AF003010; BAB53084.1; -.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000126; Ser_proteas_V8.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00839; V8PROTEASE.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 KW Complete proteome.
 SQ SEQUENCE 271 AA; 30385 MW; 37659307C63D0D9C CRC64;

Query Match 12.0%; Score 187.5; DB 16; Length 271;
 Best Local Similarity 27.6%; Pred. No. 5.8e-05;
 Matches 66; Conservative 42; Mismatches 98; Indels 33; Gaps 11;

QY 81 LKDFQTKVWIG-DGRTRKVTNTRVAPYNSIAYI--TFGG---SSCTGTLIAPNKILTNHG 134
 DB 14 LADEADYAVVPTDGRARVTHTRNRPHSVACHIERDFCDGRMTCTAFLISPTRLTLTAAH 73
 QY 135 CVYNTATR-----SYSAKGSVYFGMNDSTA-VNGSANMTEFYVPSGYINTGASQYDFAVI 188
 DB 74 CITSPIRQLGLPLNLAVKRIVTGPGRASRDARPFQGWAKQWHVNPYRRRPSGLHDVGLI 133
 QY 189 KTDNTNIGTGYRSI-----ROVTNLGT-TIKISGYPGDKMRSTGKVSOWEMSGPVTRE 242
 DB 134 ELERPFSPGFLQWSPNQDLERLNRLLHLSGTPADKPGDT---QWEHSERLDRI 189
 QY 243 DTLRLAYITDTFSGNSGSAMLQNOQ-----IVGVHNAG---YSGNGTIN---GGPKATA 290
 DB 190 TERQLFYSVDTCPCHSGAPWIIHQOAGPPWIVAVHTAGPRPHSGGAWGCRGPVPLAPA 248

RESULT 11
 Q47809 PRELIMINARY; PRT; 284 AA.
 ID AC Q47809;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE STAPHYLOCOCCAL SERINE PROTEINASE HOMOLOGUE.
 GN SPRE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OGL-10;
 RC Su Y.A., Clewell D.B.;
 RT *A gene (SprE) downstream of gelE of Enterococcus faecalis OGL-10
 RT resembles serine proteinase determinant of Staphylococcus aureus
 RT strain V8.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 RL EMBL: Z12296; CAA78168.1; -.
 DR MEROPS; S01.UBP; -.
 DR InterPro: IPR000126; Ser_proteas_V8.
 DR InterPro: IPR001254; Trypsin.
 DR PRINTS: PR00839; V8PROTEASE.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00672; V8_HIS; 1.
 DR PROSITE: PS00673; V8_SER; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 284 AA; 31063 MW; D4F0312BEE778415 CRC64;

Query Match 11.5%; Score 179.5; DB 2; Length 284;
 Best Local Similarity 25.5%; Pred. No. 0.00021;
 Matches 68; Conservative 40; Mismatches 96; Indels 63; Gaps 11;

QY 66 LSTTGEKVKPLTTLTKDQTKVVGDDGRTKVTNTRVAPYNSI-----AYITFG 115
DB 27 LSAEAEYIVPAESHSRQ-KRSLDDPDRRQEVADTTEAPFASIGRIITSPASKPGYISLG 85
QY 116 GSSCTGTLLAPKILNTHGCV----YNTATRSYSAGSV--YPCMDNSTAVNGSANMTER- 169
DB 86 ----TGFVVGTVITVNNHVAESFNKAKVNLNPNKADDAWFYGRGDSATPFGKFKVIDVA 141
QY 170 YVPSGYINTGASQYDPAVTKDTNIGNTVGYRSIR-----QVTNLTGTTI 214
DB 142 FSPNA-----DIAVV-----TVGKQNDRDPGPELGEILLTPFVLKKFESSDTHV 184
QY 215 KISGPGDKMRSTGKVSQHE-----MSGPVTREDTNLAYITDTSGSGSAMLDONQOI 269
DB 185 TISGYPGERKH-----TOWSHENDLFTSNFTDLENPLLFYDIDTTGGQSGSPIYNAQFEV 239
QY 270 VGVHNGYSGNTINGPKATAAFVEFI 296
DB 240 VGVHNGGKIQGNHGORLNEVNYNFI 266

T 12

QYKH50 PRELIMINARY; PRT; 240 AA.
AC Q9KH50;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SERINE PROTEASE SPLB.
GN SPLB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8325-4;
RX MEDLINE=21116966; PubMed=11179322;
RA Reed S.B., Wesson C.A., Liou L.E., Trumble W.R., Schlievert P.M.,
RA Bohach G.A., Bayles K.W.;
RT "Molecular Characterization of a Novel Staphylococcus aureus Serine
RT Protease Operon";
RL Infect. Immun. 69:1521-1527(2001).
DR EMBL; AF217115; AAF97926.1; -;
DR MEROPS; S01.282; -;
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00672; V8_HIS; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 240 AA; 26096 MW; 00A3349F1FE5D52E CRC64;

Query Match 11.2%; Score 174.5; DB 2; Length 240;
Best Local Similarity 28.4%; Pred. No. 0.00036;
Matches 61; Conservative 31; Mismatches 78; Indels 45; Gaps 11;
QY 78 TTYLKDFQVKVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKILNTHGCVY 137
DB 24 TTLVEEVQ-OTAKAENNVTKIQTNIFFPYTGV--VAF--KSATGFFVVGKNTILTNKHVSK 78
QY 138 N-----TATRS-----YSAGSV-YPGMNDSTAVNGSANMTEFFVPSGY-INTCA 180
DB 79 NYKVGDRITAHPNDSKNGGIYSIKKIINYPGKEDYSVQVEERAIE-RGPGFNFNDV 137
QY 181 SOYDFAVITDNTNIGNTVGYRSIRQVTNLTGTITKISGYPGDKMRSTGKVSQWEMSGPVT 240
DB 138 TPEKYAA-----GAKAGER-----IKVIGYPHP---YKNKYVLYESTGPGVM 175
QY 241 REDTNLAYITDTFSGSGSAMLDONQOQIVGVHNA 275

DB 176 SVEGSSIVYSAHTESGSGSPVLNSNNELVGIHFA 210
RESULT 13
Q53781 PRELIMINARY; PRT; 240 AA.
AC Q53781;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NOVEL ANTIGEN (SERINE PROTEASE SPLB).
GN SPLB OR SA1630 OR SAV1812.
OS Staphylococcus aureus, (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280, 158879, 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=FDA #485 (ABD ENT+);
RA Rieneck K., Renneberg J., Gutschik E., Diamant M., Svenson M.,
RA Bendtzen K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
CC -! SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; U60589; AAB03613.1; -;
DR EMBL; AP003135; BAB42898.1; -;
DR EMBL; AP003363; BAB57974.1; -;
DR MEROPS; S01.282; -;
DR InterPro; IPR000126; Ser_proteas_v8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00672; V8_HIS; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 240 AA; 26140 MW; 45EF0B5A51E05DA7 CRC64;

Query Match 11.1%; Score 173.5; DB 16; Length 240;
Best Local Similarity 27.9%; Pred. No. 0.00042;
Matches 60; Conservative 32; Mismatches 78; Indels 45; Gaps 11;
QY 78 TTYLKDFQVKVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKILNTHGCVY 137
DB 24 TTLVEEVQ-OTAKAENNVTKIQTNIFFPYTGV--VAF--KSATGFFVVGKNTILTNKHVSK 78
QY 138 N-----TATRS-----YSAGSV-YPGMNDSTAVNGSANMTEFFVPSGY-INTCA 180
DB 79 NYKVGDRITAHPNDSKNGGIYSIKKIINYPGKEDYSVQVEERAIE-RGPGFNFNDV 137
QY 181 SOYDFAVITDNTNIGNTVGYRSIRQVTNLTGTITKISGYPGDKMRSTGKVSQWEMSGPVT 240
DB 138 TPEKYAA-----GAKAGER-----IKVIGYPHP---YKNKYVLYESTGPGVM 175
QY 241 REDTNLAYITDTFSGSGSAMLDONQOQIVGVHNA 275
DB 176 SVEGSSIVYSAHTESGSGSPVLNSNNELVGIHFA 210


```

RT *Molecular Characterization of a Novel Staphylococcus aureus Serine
RT Protease Operon.;
RL Infect. Immun. 69:1521-1527(2001).
DR EMBL; AF271715; AAF97927.1; -
DR MEROPS; S01.283; -
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR000126; Ser_proteas_V8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00672; V8_HIS; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 239 AA; 26098 MW; BA797E1478652DBA CRC64;

Query Match 10.0%; Score 156; DB 2; Length 239;
Best Local Similarity 25.9%; pred. No. 0.0058;
Matches 57; Conservative 29; Mismatches 88; Indels 46; Gaps 11;

QY 96 TKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKLITNGHCY--YNTATP----- 142
DB 41 TVQKDTNIFYNGV--VSF--KDATGFVIGNTIITNKHVSKDYKVGDRITAHNGDKGN 96

QY 143 --SYSAKG-SVYPCMDNSTAVNGSANMTEFVYPSGY-INTGASQYDFAVIKTDNIGNTV 198
DB 97 GGIYKIKSISDYPGDEDISVMNIEEQAVE-RGPKGFNFENVQAFNFA---KDAKVDD-- 150

QY 199 GYRSIROVTNLTTGTTIKISYPGDKMSTGKVSQWENSGPVYTRDTNLAYTYITDFFSGNS 258
DB 151 -----KIKVIGYP--LPAQNSFKQFESTGTTIKRIKDNILNFDAYIEPGNS 193

QY 259 GSAMLDNQOQIVGVHNNAGYS--NCTINGGPKRTAAAFVEFI 296
DB 194 GSPVLNSNNEVIGVVGIGIKIGSEYNGAVYFTPIQKDFI 233

Search completed: June 21, 2002, 18:45:41
Job time: 220 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:40:36 ; Search time 17.42 Seconds
(without alignments)
1665.843 Million cell updates/sec

Title: US-09-652-743A-12
Perfect score: 1559
Sequence: 1 MKVKRLIPSLLVFGALSVP.....NGCPKATAAFVEFINAKAQ 302

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	420.5	27.0	316	2 A45134	endopeptidase (EC
2	280.5	18.0	313	2 A35122	metalloproteinase
3	202.5	13.0	188	2 E90065	hypothetical prote
4	192.5	12.3	357	2 S21758	glutamic acid-spec
5	189.5	12.2	336	1 PRSASK	glutamyl endopepti
6	189.5	12.2	342	2 G89873	hypothetical prote
7	182.5	11.7	278	2 AH0282	probable peptidas
8	179.5	11.5	284	2 S25140	serine proteinase
9	173.5	11.1	240	2 C89967	serine proteinase
10	170.5	10.9	218	2 E97915	choline binding pr
11	156	10.0	330	2 B82415	probable serine pr
12	155.5	10.0	280	1 PRSAEA	epidermolytic toxi
13	154	9.9	273	2 B85765	hypothetical prote
14	153.5	9.8	273	2 H64915	putative protease
15	153.5	9.8	285	2 C95045	choline binding pr
16	150	9.6	239	2 B89967	serine proteinase
17	141.5	9.1	239	2 H89966	serine proteinase
18	139.5	8.9	235	2 D89967	serine proteinase
19	137.5	8.8	239	2 A89967	serine proteinase
20	135	8.7	273	2 A80682	probable secreted
21	133.5	8.6	168	2 D90066	hypothetical prote
22	133	8.5	277	1 PRSAEB	epidermolytic toxi
23	130	8.3	1275	2 T33369	hypothetical prote
24	128.5	8.2	272	2 JC4170	trypsin-like prote
25	125	8.0	1651	2 JC1340	outer membrane pro
26	124	8.0	698	2 A59222	hypothetical prote
27	121	7.8	397	2 H70789	probable serine pr
28	120.5	7.7	488	2 B69825	cell wall-binding
29	120	7.7	254	1 TRWV3Y	trypsin-like prote

ALIGNMENTS

RESULT 1

A45134
endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A45134; S23078
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, I.
J. Biol. Chem. 267, 23782-23788, 1992
A:Title: Purification, characterization, cloning, and expression of a glutamic ac
A:Reference number: A45134; MUID:93054737
A:Accession: A45134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KAK>
A:Cross-references: GB:D10060; NID:g216263; PIDN:RAA00949.1; PID:d1001415; PID:g2
A:Experimental source: ATCC 14580
A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIIP:118785)
R:Svendsen, I.; Breddam, K.
Eur. J. Biochem. 204, 165-171, 1992
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptic
A:Reference number: S23078; MUID:92155199
A:Accession: S23078
A:Status: preliminary
A:Molecule type: protein
A:Residues: 95-316 <SVE>
C:Keywords: hydrolase

Query Match 27.0% ; Score 420.5; DB 2; Length 316;
Best Local Similarity 33.0% ; Pred. No. 2.le-22;
Matches 101; Conservative 55; Mismatches 121; Indels 29; Gaps 10;

Qy 5 KKLIPSLLVFGALSVPSPA---HAASDVLTSVDYDMVTSVG---KVISSADFNHDKMT--- 56

Db 4 KSKVKRGLITGLIGISYSLGMHPA--QAAPSPHTPVSDDPSYKRAETSVTDPNKISDQY 61

Qy 57 ---PSSFPKVDLSTICEKVKPLTYLKDFQTKVIGDGGRTKVTNTRVAPYNSIAVIT 113

Db 62 GLYSKAFGTGKVNKT-KERAEEKSPAKAPYSIKSVIGSDDRTRVTNTATPYRAIVHIS 120

Qy 114 FGGSSCTGLIAPKILTNHGVNTATRSYSAKGSVPGMNDSTAVNGSANMTFFYVPS 173

Db 121 SSIGSCTGWMIGPKVTATAGHCIVDTSSGSFAGTATVSPGRNGTSYPGSVKSTFYFIPS 180

Qy 174 GYINTGASQYDPAVIKTDNTNIGNTVGYRSIRQVT-NLTGTTIKISGYPGDKMRSTGKVSQ 232

Db 181 GW-RSGNTNYDGAITELSEPICNTVGYFGYSYTTSSLVGTTVTSIGYPGDKTAGT---Q 235

Qy 233 WEMSGPVTREDTNLAYTYTIDTFSGNSGSAMLDMQNOQ-----IVGVHNAG-YSNCTIN 283

Db 236 WOHSGPIAISETYKLQYAMDTYGGSGSPVFEEQSSSRNTNCSGPCSLAVHTNGVYGGSSYN 295

Db 170 YPYGTYSGTTFPYSVKGVWTEKSDNTDYGAIKLNGSPGNTGVGYGRTTNSSPVGLSSV 229
 QY 217 SGYPGDKMRSTGKVSQWEMSGPVTRDTNLAYYYTIDTFSGNSGSAML-----DQNOQIVGV 272
 Db 230 TGPDCDKTGT-----MMSDTPKPIRSAETYLKLYTTDRYGCOSGSPVYRNYSDTGOTAI 285
 QY 273 HNAGYSNGTINGGPKATAAFVEFNY 298
 Db 286 HTNGGS--SYNLGTRVTDNVFNIOY 309
 RESULT 3
 E90065
 hypothetical protein SA2382 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: E90065
 R.;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: E90065
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <KUR>
 A:Cross-references: GB:BA000018; PID:g13702546; PIDN:BA043687.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA2382
 Query Match 13.0%; Score 202.5; DB 2; Length 188;
 Best Local Similarity 40.4%; Pred. No. 2.4e-07;
 Matches 46; Conservative 22; Mismatches 37; Indels 9; Gaps 4;
 QY 186 AVIKTDNIGNTVGRSIRQVNTLT-GTTIKISYPGDKMRSTGKVSQWEMSGPVTRDT 244
 Db 2 AIKLNLSNLGNKTYGLTLN--THISKGENEISGFGDKSDN----ROYKKGKLESFDE 55
 QY 245 NLAYYYTIDTFSGNSGSAMLQDNOQIVGVHNAGYSNGTINGGPKATAAFVEFNY 298
 Db 56 NEMYTYVDTFSGGSAIRDSKNNIIGVHAYGRYNH--NSGVRINDLKLDYINY 107
 RESULT 4
 S21758
 glutamic acid-specific endopeptidase - Staphylococcus aureus
 C:Species: Staphylococcus aureus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S21758
 R.;Yoshikawa, K.; Tsuzuki, H.; Fujiwara, T.; Nakamura, E.; Iwamoto, H.; Matsumoto, Biochim. Biophys. Acta 1121, 221-228, 1992
 A:Title: Purification, characterization and gene cloning of a novel glutamic acid-
 A:Reference number: S21758; MUID:92287954
 A:Accession: S21758
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-357 <YOS>
 A:Cross-references: GB:D00730; NID:g216970; PIDN:BAA00630.1; PID:g216971
 C:Superfamily: staphylococcal serine proteinase
 Query Match 12.3%; Score 192.5; DB 2; Length 357;
 Best Local Similarity 23.8%; Pred. No. 2.9e-06;
 Matches 77; Conservative 47; Mismatches 127; Indels 73; Gaps 14;
 QY 3 KVKKLIPLSLVFGALSVPSFAHAASDVLSYSDYDMVTSDGKVISSADFHNDKMTFSPFDK 62
 Db 7 KVSSLFVATLTTATLVSSPAANALSSKAMDHNHPOQTQTD-----KQQTTP-----K 51

Db 170 YPYGTYSGTTFPYSVKGVWTEKSDNTDYGAIKLNGSPGNTGVGYGRTTNSSPVGLSSV 229
 QY 217 SGYPGDKMRSTGKVSQWEMSGPVTRDTNLAAYITIDTFSGNSGSAML-----DQNOQIVGV 272
 Db 230 TGPDCDKTGT-----MMSDTKPIRSAETYLKLTITTDYRGCGSGSPVYRNYSDTGOTAI 285
 QY 273 HNAGYSNGTINGGPKATAAFVEFNY 298
 Db 286 HTNGGS--SYNLGTRVTDNVFNIOY 309
 RESULT 3
 E90065
 hypothetical protein SA2382 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: E90065
 R.;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: E90065
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <KUR>
 A:Cross-references: GB:BA000018; PID:g13702546; PIDN:BA043687.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA2382
 Query Match 13.0%; Score 202.5; DB 2; Length 188;
 Best Local Similarity 40.4%; Pred. No. 2.4e-07;
 Matches 46; Conservative 22; Mismatches 37; Indels 9; Gaps 4;
 QY 186 AVIKTDNIGNTVGRSIROVTNLT-GTTIKISYPGDKMRSTGKVSQWEMSGPVTRDT 244
 Db 2 AIKLNLSNLGNKTYGLTN--THISKGENEISGFGDKSDN----ROYKKGKLESFDE 55
 QY 245 NLAAYITIDTFSGNSGSAMLQDNOQIVGVHNAGYSNGTINGGPKATAAFVEFNY 298
 Db 56 NEMYTYDVTSGGSGAIRDKNKIIGVHAYGRYNH--NSGVRINDLKLDYINY 107
 RESULT 4
 S21758
 glutamic acid-specific endopeptidase - Staphylococcus aureus
 C:Species: Staphylococcus aureus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S21758
 R.;Yoshikawa, K.; Tsuzuki, H.; Fujiwara, T.; Nakamura, E.; Iwamoto, H.; Matsumoto, Biochim. Biophys. Acta 1121, 221-228, 1992
 A:Title: Purification, characterization and gene cloning of a novel glutamic acid-
 A:Reference number: S21758; MUID:92287954
 A:Accession: S21758
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-357 <YOS>
 A:Cross-references: GB:D00730; NID:g216970; PIDN:BAA00630.1; PID:g216971
 C:Superfamily: staphylococcal serine proteinase
 Query Match 12.3%; Score 192.5; DB 2; Length 357;
 Best Local Similarity 23.8%; Pred. No. 2.9e-06;
 Matches 77; Conservative 47; Mismatches 127; Indels 73; Gaps 14;
 QY 3 KVKKLIPSLIFVFGALSVPSFAHAASDVLSYSDYDMVTSDGKVISSADFHNDKMTFSSFDK 62
 Db 7 KVSSLFVATLTATLVSSPAANALSSKAMDHNHPOQTQTD-----KQQTTP-----K 51

RESULT 6

hypothetical protein sspA [Imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G89873
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.;
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus* strain N315
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: G89873
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-342 <KUR>
A;Cross-references: GB:BA000018; PID:gl3700850; PIDN:BA642146.1; GSPDB:GN000001
A;Experimental source: strain N315

C; Genetics: A; Gene: *sspA*
C; Superfamily: staphylococcal serine proteinase

Db	7	KVSSLFVATLTATLVSSPAANALS-----SKAMDNHPQQTSSKQ	48
Qy	63	VDDLSSTIGEKKVKPLTTYLKDFQKWIGDGRTKVTNTRVAPYNSIAYITF----GGSS	118

```

QY 226 GRVSQWMSGPFVREDINRLIILDIFGNSGSGMADQQAQQAQVGNHNSGIRGIIINGRKA 201
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 ----MWESKGTKITLKGEMQYDLSTTGGNSGSPVFNEKNEYIGIHWGVPN-EFNG--- 262

```

QY	200	AAAFV	293
		:	
Db	263	--AVFI	266

RESULT 7
AH0282

C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AF0282

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead,
Nature 413, 523-527, 2001

Db 265 ----VGIQVVK 272

RESULT 13

E85765

hypothetical protein 22592 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: E85765

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Martinez, D.; Dickson, M.C.; Karp, P.; Lator, R.; Liu, X.; Nolte, H.; Paulsen, O.; Saxe, J.; Shao, Y.; Sun, J.; Taylor, T.; Tiedje, J.M.; White, O.; Young, K.; Zeng, A.P.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, L.; Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <STO>

A:Cross-references: GB:AE005174; NID:g12515576; PIDN:AA056585.1; GSPDB:GN00145; UWGP:22592

A:Experimental source: strain O157:H7, substrain EDL933

A:Keywords: e: 22592

Query Match 9.8%; Score 154; DB 2; Length 273;

Best Local Similarity 23.6%; Pred. No. 0.001;

Matches 57; Conservative 40; Mismatches 91; Indels 54; Gaps 10;

QY 65 DLSSTIGERKVPKLTTLKDFQTKWIGDGRKVTNTRVAPYNSIAYI-TFGSSCTGTL 123

Db 25 DVAKSANDEVSL-----FFGHDRVPVNDTQSPWDVAGQLETASGNLCTATL 73

QY 124 IAPNKLITNGHCYVN-----TATRSYSAK-----GSVYPGMNDSTAVNGS 163

Db 74 IAPNLAHTAGHCLLTPPKGKADKAVLRVSNKGLWRYDIHDIIEGRVDP TLCKRLKADGD 133

QY 164 ANMTEFYPVSGYINTGASQYDFA-VIKTDTNIGNTV-----GYRSIROVTNLTGTTIKI 216

Db 134 G-----WIVP-----PAAQPMDFGLVLRNPPSGITPLPLEFGDKAALTAALKAAGRKVTQ 184

QY 217 SGYPGDKMRSTGRVSGWEMSGPVTRDTNLAYTTIDTFSGNSGSAML-----DONQOIVGVH 273

Db 185 AGYPEDHLDTLHSHQNCVETG---WAQTSVMHQCDTLPDGSGLPLMLHTDDGWLIGVQ 241

QY 274 NA 275

Db 242 SS 243

RESULT 14

active protease ydgD precursor (EC 3.4.21.-) [similarity] - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 31-Dec-2001

C:Accession: H64915

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H64915

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <BLAT>

A:Cross-references: GB:AE000255; GB:U00096; NID:g1787875; PIDN:AA074670.1; PID:g1787882;

A:Experimental source: strain K-12, substrain MG1655

A:Keywords: hydrolase; serine proteinase

Query Match 9.8%; Score 153.5; DB 2; Length 273;

Best Local Similarity 24.6%; Pred. No. 0.001;

Matches 55; Conservative 37; Mismatches 89; Indels 43; Gaps 9;

QY 83 DFQTKWVIGDGRKVTNTRVAPYNSIAYI-TFGSSCTGTLIAPNKLITNGHCYVN---- 138

Db 32 DEVSTLFFGHDRVPVNDTQSPWDVAGQLETASGNLCTATLAPNLAHTAGHCLLTPPK 91

QY 139 -----TATRSYSAK-----GSVYPGMNDSTAVNGSANMTEFYPVSGYINTGAS 181

Db 92 GRADKAVALLRFVSNKGLWRYDIHDIIEGRVDP TLCKRLKADGD-----WIVP-----PAAA 142

QY 182 QYDFA-VIKTDTNIGNTV-----GYRSIROVTNLTGTTIKISGYPGDKMRSTGRVSGW 234

Db 143 PWFDFGLVLRNPPSGITPLPLEFGDKAALTAALKAAGRKVTQAGYPEDHLDTLHSHQNC 202

QY 235 MSSPVTRDTNLAYTTIDTFSGNSGSAML-----DONQOIVGVHNA 275

Db 203 VTG---WAQTSVMHQCDTLPDGSGLPLMLHTDDGWLIGVSS 243

RESULT 15

C95045

choline binding protein G [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: C95045

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; On, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtz, T.; Hickey, E.K.; Holt, I.E.

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: C95045

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-285 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74556.1; PID:g14971860; GSPDB:GN00164; TIGR4

A:Experimental source: strain TIGR4

A:Genetics: C:Gene: SP0390

Query Match 9.8%; Score 153.5; DB 2; Length 285;

Best Local Similarity 27.0%; Pred. No. 0.0012;

Matches 50; Conservative 26; Mismatches 94; Indels 15; Gaps 6;

QY 120 TGTLIAPNKLITNGHCYVNTRATRSYSAGS-----VYPMNDSTAVNGSANMTEFYPVSGYI 176

Db 20 SANFIKDNVLIATAAHNYR---HDYKKEADDIYVLPVAVSPSOEPFGKIKVKEVYLKEFR 76

QY 177 NTG---ASQYDFAVIKTDTNIGNTVGYSI-RQVTNLTGTTIKISGYPGDKMRSTGRVSG 232

Db 77 NLNSKDAREYDLALLILEPIGAKLTGLTSLPTSQKLTGTTITGTPSYNF-----KIHO 132

QY 233 -WEMSGPVTRDTNLAYTTIDTFSGNSGSAMLDOQOIVGVHNAVYSGTNGGPKATAA 291

Db 133 MYTDKKQVLSDDGMFLDYQVDTLEGSSGTVYDASHRVVGVHTLGDGANOINSVAKLNER 192

QY 292 FVEFI 296

Db 193 NLFFI 197

Search completed: June 21, 2002, 18:42:41

Job time: 125 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:42:21 ; Search time 13.5 Seconds
(without alignments)
866.171 Million cell updates/sec

Title: US-09-652-743A-12
Perfect score: 1559
Sequence: 1 MKKVKLIPSLVFGALSVP.....NGGPKATAAFVEFINAKAQ 302

ing table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	420.5	27.0	316	1 GSEP_BACLI	P80057 bacillus li
2	280.5	18.0	313	1 MPR_BACSU	P39790 bacillus su
3	189.5	12.2	336	1 STSP_STAUA	P04188 staphylococ
4	153.5	10.0	280	1 ETA_STAUA	P09331 staphylococ
5	133	8.5	273	1 YDGD_ECOLI	P76176 escherichia
6	123	8.5	277	1 ETB_STAUA	P09332 staphylococ
7	123	7.9	1656	1 OMPB_RICJA	O06653 r outer mem
8	120	7.7	254	1 TRY3_AEDAE	P29786 aedes aegypt
9	119	7.6	678	1 YF48_MYCTU	Q10778 mycobacteri
10	118.5	7.6	227	1 TRYP_SACER	P24664 saccharopol
11	118	7.6	248	1 TRYP_FUSOX	P35049 fusarium ox
12	116.5	7.5	389	1 SERI_BOMMO	P07856 bombyx mori
13	116	7.4	280	1 TRY2_DROME	P42280 drosophila
14	113.5	7.3	256	1 HYPA_HYPLI	P35587 hypoderma l
15	113	7.2	569	1 YE59_MYCPN	P75324 mycoplasma
16	112	7.2	281	1 TRY2_DROER	P54630 drosophila
17	112	7.2	679	1 PAN3_YEAST	P36102 saccharomyc
18	111	7.1	1286	1 AIDA_ECOLI	Q03155 escherichia
19	110	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
20	109	7.0	264	1 GRAK_HUMAN	P49863 homo sapien
21	109	7.0	532	1 SPG7_DICDI	P22698 dictyosteli
22	109	7.0	762	1 SLAP_ACEKI	P22258 acetogenium
23	108.5	7.0	1300	1 L20K_RICRI	P14914 rickettsia
24	108.5	7.0	1654	1 OMPB_RICPI	Q53047 r outer mem
25	108	6.9	262	1 TRY2_DROME	P42278 drosophila
26	108	6.9	1169	1 YR82_YEAST	P36170 saccharomyc
27	107.5	6.9	214	1 ACH2_LONAC	P23605 lonomia ach
28	107.5	6.9	348	1 ASG2_ECOLI	P00805 escherichia
29	107.5	6.9	1250	1 YFAL_ECOLI	P45508 escherichia
30	107	6.9	636	1 YNR6_YEAST	P53882 saccharomyc
31	106.5	6.8	497	1 FLIC_ECOLI	P04949 escherichia
32	106	6.8	664	1 Y4FB_RHISN	P55440 rhizobium s
33	106	6.8	2249	1 OMPA_RICRI	P15921 rickettsia

RESULT 1

ID	GSEP_BACLI	STANDARD	PRT	316 AA
AC	P80057			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Glutamyl endopeptidase precursor (EC 3.4.21.19) (Glutamate specific endopeptidase) (GSE).			
DE	BLASE.			
GN	Bacillus licheniformis.			
OS	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1402;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=ATCC 14580;			
RX	MEDLINE=93054737; PubMed=1429718;			
RA	Kakugo S, Kikuchi N, Kitadokoro K, Fujiwara T., Nakamura E., Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;			
RA	*Purification, characterization, cloning, and expression of a glutamic acid-specific protease from Bacillus licheniformis ATCC 14580.*;			
RT	J. Biol. Chem. 267:23782-23788(1992).			
RL	[2]			
RN	SEQUENCE OF 95-316.			
RP	MEDLINE=92155199; PubMed=1346764;			
RX	Svendsen I., Bredam K.;			
RA	Isolation and amino acid sequence of a glutamic acid specific endopeptidase from Bacillus licheniformis.*;			
RT	Eur. J. Biochem. 204:165-171(1992).			
RL	-1- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU.			
CC	-1- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.			
CC	-1- SUBCELLULAR LOCATION: Secreted			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D10060; BAA00949.1; -;
PIR: S23078; S23078.
PIR: A45134; A45134.
MEROPS: S01.271; -;
InterPro: IPR000126; Ser_proteas_v8.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; trypsin; 1.
PRINTS: PM00839; V8PROTEASE.
SMART: SM00020; Tryp_Spc; 1.

ALIGNMENTS

34	105	6.7	272	1	SER3_DROME	P17207 drosophila
35	105	6.7	653	1	API_ACHLY	P15636 achromobact
36	105	6.7	928	1	PM11_CHLPN	O86164 chlamydia p
37	105	6.7	1005	1	Y321_MYCPN	P75327 mycoplasma
38	103.5	6.6	686	1	MAS2_HUMAN	O00187 homo sapien
39	103	6.6	1041	1	EGT2_YEAST	P42835 saccharomyc
40	102.5	6.6	258	1	VSP1_TRIGA	O13059 trimeresuru
41	101.5	6.5	1953	1	BIGA_SALTY	P25927 salmonella
42	101	6.5	261	1	EUM3_EURMA	O97370 euroglyphus
43	101	6.5	265	1	SER1_DROME	P17205 drosophila
44	101	6.5	1150	1	APMU_PIG	P12021 sus scrofa
45	100.5	6.4	1655	1	OMPB_RICCN	Q9AKA3 r outer mem


```
DR PROSITE: PS00672; V8_HIS: 1.
DR KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 94
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 33611 MW; 96D7552CB7089B09 CRC64;

Query Match 27.0%; Score 420.5; DB 1; Length 316;
Best Local Similarity 33.0%; Pred. No. 4.1e-23;
Matches 101; Conservative 55; Mismatches 121; Indels 29; Gaps 10;

QY 5 KKLIPSLVPGALSPVFA--HAASDVLTSYDMVTSDG--KVISSADFHNDKMT---56
R 4 KRSVKRLITGLIGISYLGMPA--QAAPSPHTPVSSDPSYKAETSVTYDPNKSQY 61

b 57 ---PSSFQKDDLSSTIGERKVKPLTYLKDFTQKVVIGDGRKVTNTRVAPYNSIAYIT 113
62 GLYSKAFPTGKYNET-KEAEKSPAKAPYSIKSVIGSDRTRVNTTAYPYRAIVHIS 120

QY 114 FGSSCTGTLLIANKILTNHCHVYNTATRSYSAKGSYVPGMNDSTAVNGSANTEFYVPS 173
Db 121 SSIGSCTGWMIGPKTVATAGHCYDTSSTGSGFAGTATVSGRNGTSYPYGSVKSTRYFIPS 180

QY 174 GYINTGASQYDFAVIKRTDNTGNVTGYRSIROVT-NLTGTTIKISYPGDKMRSTCKVSO 232
Db 181 GW-RSGNTNDYGAIELEPISGNTVGTGYSYTTSLVGLTITVTSIGPGDKTAGT---Q 235

QY 233 WEMSGPVTRDTNLYITIDTFSGNSGSAMLDQNOQ-----IVGVHNAG-YSNGTIN 283
Db 236 WOHSGPIAISETKLYQYAMDYTGQSGSPVFEGSSSRNCSGCSLAVHTNGYVGSSYN 295

QY 284 GGPAT 289
Db 296 RGRIT 301

RESULT 2
MPR_BACSU STANDARD; PRT; 313 AA.
AC P39790;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extracellular metalloprotease precursor (EC.3.4.21.-).
GN MPR.
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=1423;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-GP241;
RX MEDLINE-90130256; PubMed-2105291;
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A.,
RA Ally D., Pero J.;
RT "Gene encoding a novel extracellular metalloprotease in Bacillus
subtilis."
RL J. Bacteriol. 172:1024-1029(1990).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-168;
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
Bacillus subtilis chromosome."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 1-68 FROM N.A.
```

```
RX MEDLINE-89108019; PubMed-3145906;
RA Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
Bacillus subtilis chromosome."
RL Gene 70:351-361(1988).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN-168;
RX MEDLINE-96118702; PubMed-7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purt gene encoding
formate-dependent glycinamide ribonucleotide transferase."
RL Microbiology 141:2211-2218(1995).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
V8 FAMILY.
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2
FAMILY OF SERINE PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10505; AAA22604.1;
DR EMBL; AB006424; BAA33121.1;
DR EMBL; M22916; AAA22832.1;
DR EMBL; Z99105; CAB12018.1;
DR PIR; A35122; A35122.
DR HSSP; P00756; ISGF.
DR MEROPS; S01.272;
DR Subtilist; BG10690; mpr.
DR InterPro; IPR000126; Ser.proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 34 POTENTIAL.
FT PROPEP 35 93
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 131 147 BY SIMILARITY.
FT CONFLICT 61 68 QVSAPYEG -> PLESTAOA (IN REF. 3).
SQ SEQUENCE 313 AA; 33842 MW; D41788E8D652AE94 CRC64;

Query Match 18.0%; Score 280.5; DB 1; Length 313;
Best Local Similarity 29.4%; Pred. No. 3.6e-13;
Matches 96; Conservative 50; Mismatches 129; Indels 51; Gaps 15;

QY 6 KLIPSLV--FGALSPVFAHAASD-----VLTSDYDMVTSDGKVISSADPHNDKMT 57
Db 2 KLVPRFRKQFAVLTVLCLALAAAVSGVPAKAAENPQTSVSTGK---EAD-----ATK 53
QY 58 SSFDKVDLISST---IGEKVKPL---TTLKDFQT---KVVIGDGRKVTNTRVAPYN 107
Db 54 NOTSKADQVSAPYEGTCKTSKSLYGGQTELEKNIOFPSSIIITGDTERTISSTTFPYR 113
QY 108 -----SIAY-----ITFGSSCTGTLLIAPNKILTNHCHVYNTATRSYSAKGSVYPGMNDST 158
Db 114 ATVOLSIKPTNTSTVG---CTGLVNPNTVITAGHCYV-SQDHGWASTITAAPGRNGSS 169
QY 159 AVNGSANMTEFYVPSGYINTGASQYDFAVIKRTDNTGNVTGYRSIROVTNLT--GTTIKI 216
Db 170 YPYGTYSGMTFYSVKGWGTESKDTNIDYGAIKLNGSPGNTVGVGYRTNNTSSSPVGLSSV 229
```

QY 217 SGYPGKMRSTGKVSQWEMSGPVTRDTNLAAYTIDTFSCNSGSAI-----DQNOQIVGV 272
DB 230 TGPCKDTGT-----MMSDTKPIRSAETKYLTYTDTYGCQSGPVRYRNSYTGQTAIAI 285
QY 273 HNAGYSNGTINGCKPATAAFVEFINY 298
DB 286 HTNGGS--SYNLGTRVTDVFNNOY 309

RESULT 3

ID STSP_STAAU STANDARD; PRT; 336 AA.
AC P04188:
DT 20-MAR-1987 (Rel. 04, Created)
01-APR-1988 (Rel. 07, Last sequence update)
01-MAR-2002 (Rel. 41, Last annotation update)
Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
proteinase) (V8 proteinase) (Endoproteinase Glu-C).
SSPA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-V8;
RX MEDLINE=87316953; PubMed=3306605;
RA Carmona C., Gray G.L.;
RT "Nucleotide sequence of the serine protease gene of Staphylococcus
aureus, strain V8."
RL Nucleic Acids Res. 15:6757-6757(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20569178; PubMed=11119502;
RA Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;
RT "Description of staphylococcus serine protease (ssp) operon in
Staphylococcus aureus and nonpolar inactivation of sspA-encoded serine
protease."
RL Infect. Immun. 69:159-169(2001).
RN [3]
RP SEQUENCE OF 69-280.
RX STRAIN-V8;
RX MEDLINE=78212487; PubMed=969922;
RA Drapeau G.R.;

The primary structure of staphylococcal protease.;

Can. J. Biochem. 56:534-544(1978).

-1- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-
TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.

-1- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
V8 FAMILY.

-1- DATABASE: NAME=Worthington enzyme manual;

WWW="http://www.worthington-biochem.com/manual/P/STAP.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; Y00356; CAA68434.1; .

DR EMBL; AF309515; AAG45843.1; .

DR PIR; A26812; PRSASK.

DR MEROPS; S01.269; .

DR InterPro; IPR000126; Ser_proteas_v8.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00839; V8PROTEASE.

DR PROSITE; PS00672; V8_HIS; 1.

DR PROSITE; PS00673; V8_SER; 1.

DR Hydrolase; Serine protease; Zymogen; Signal.

FT SIGNAL 1 29
FT PROPEP 30 68
FT CHAIN 69 336
FT ACT_SITE 119 119
FT ACT_SITE 161 161
FT ACT_SITE 237 237
FT CONFLICT 109 109
FT CONFLICT 125 125
FT CONFLICT 145 145
FT CONFLICT 193 193
FT CONFLICT 229 229
FT CONFLICT 259 261
FT CONFLICT 268 270
SQ SEQUENCE 336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;

Query Match 12.2%; Score 189.5; DB 1; Length 336;

Best Local Similarity 24.2%; Pred. No. 1.2e-06;

Matches 74; Conservative 45; Mismatches 126; Indels 61; Gaps 12;

QY 3 KYKKLIPSLLVFGALSVPSFAHAASVLTSDYDMVTSQKVISSADFHNDMKTSPSFDK 62
DB 7 KVSLLFVATLTATLVSSPAANALS-----SKAMDNHPQQTQSSKQ 48
QY 63 VDDLSTIGEKKVPLTYLKDFQTKVIGDDGRTKVTNTRVAPYNSIAYITF-----GGSS 118
DB 49 TPRIQK--GGNLKPLE---QREHANVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTFI 103
QY 119 CTGTLIAPNKLITNGHCYVNTATRSYSAGSVYP-CMNDSTAVNGSANMTEFVVPQSYIN 177
DB 104 ASGVVGVKDTLLTNKHVDATHGDPHALK--APPSAINQDNPNGGFTAEQITKYSG--- 158
QY 178 TGASQYDFAVIKTDTNIGN-----TVGYRSIRSQVTNLTGTTIKISGYPGDKMRST 227
DB 159 ----ECGLAIVKFSPEONKHGCVVVKPATMSNAETQVNO---NITVGYGCDKPVAT 210
QY 228 GKVSQWEMSGPVTRDTNLAAYTIDTFSCNSGSAI-----DQNOQIVGVNAGYSNGTINGGPK 287
DB 211 ----MWESKGRITLYLKGEMAHQYDLSTGTGNSGSPVFNKNEKVIHGWGVPN-EFNG--- 262
QY 288 ATAAFY 293
DB 263 --AVFI 266

RESULT 4

ETA_STAAU STANDARD; PRT; 280 AA.
AC P09331;
DT 01-MAR-1989 (Rel. 10, Created)
01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exfoliative toxin A precursor (EC 3.4.21.-) (Epidermolytic toxin A).
GN ETA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-UT0002;
RX MEDLINE=87307959; PubMed=3040666;
RA Lee C.Y., Schmidt J.J., Johnson-Winegar A.D., Spero L., Iandolo J.J.;
RT "Sequence determination and comparison of the exfoliative toxin A and
toxin B genes from Staphylococcus aureus."
RL J. Bacteriol. 169:3904-3909(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TC16;
RX MEDLINE=87307960; PubMed=3040667;
RA O'Toole P.W., Foster T.J.;
RT "Nucleotide sequence of the epidermolytic toxin A gene of
Staphylococcus aureus."

Pfam: PF00089; trypsin: 1.
 PRINTS: PR00839; V8PROTEASE.
 DDR PROSITE: PS00672; V8_HIS: 1.
 PROSITE: PS00673; V8_SER: 1.
 KKW Hydrolase; Serine protease; Calcium; Toxin; Signal; 3D-structure.
 SIGNAL 1 38
 CHAIN 39 280
 ACT_SITE 110 110
 ACT_SITE 158 158
 ACT_SITE 233 233
 MUTAGEN 233 233
 S->G: LOSS OF TOXICITY.
 SEQUENCE 280 AA; 31077 MW; 4FBAF750FFE43586 CRC64;

Query Match 10.0%; Score 155.5; DB 1; Length 280;
 Best Local Similarity 24.7%; Pred. NO. 0.00025;
 Matches 77; Conservative 46; Mismatches 132; Indels 57; Gaps 13;

QY 2 KVKKKLIPSLVFGALSVPSFAHAASDVLTSYDMV-TSDGKVISSADPHNDMKTPSSF 60
 DDB 5 KIISKVLLSLSF---TVGASAFVIODELQMKHKAESAEEIKHKEEKWKNYYGVNAF 61
 QY 61 DKVDDLSSTIGEEKVKPLTTLTKDFQTKVIGDDGRTKVTRVAPVNSIAYI-TFGSSC 119
 DDB 62 NLPKELFSKDEK-----DRQKPYNTIGNVFPVKQTS 95
 QY 120 TGTLIAPNKILTNHGCVYNTATRSYSAKGSVYPCMDNSTAVNGSANMTEFYVPSGYINTG 179
 DDB 96 TGVVLGKNTVLTNRHIA--KFANGDFSKVSFRFSINTDDNGNTETPYGEYEVKEILQEPF 153
 QY 180 ASQYDPAVT--KTDNTN---IGNTVGYRSTRQVTNL-TGTTIKISGYPGDKMRSTGKVSQW 233
 DDB 154 GAGVDLALIRLPDQNGVSLGDKISPAKIGTSDNLKDGDKLELIGYFPDH-----KYNQM 208
 QY 234 ENSG-PVTREDTNLAYTITDTSNGSSGSAMLDQNOQITGVVHNAGYSN---GTINGPKA 288
 DDB 209 HRSEIELTTLRGLRYIGF-TVPCNSGSGIFNSNGELVGIHSSKVSCHLDREHQINYG--- 264
 QY 289 TAAFVEFINTAK 300
 DDB 265 ----VGIGNYVK 272

RESULT 5
 YDGD_ECOLI STANDARD; PRT; 273 AA.
 AC P76176:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative protease ydgd precursor (BC 3.4.21.-).
 GN YDGD OR B1598.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid:562;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.C., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
 CC V8 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

RX MEDLINE=90351379; PubMed=2117445;
RT Bailey C.J., Smith T.P.;
RA "The reactive serine residue of epidermolytic toxin A";
RL Blochem. J. 269:535-537(1990).

[4]	FUNCTION.
CC	MEDLINE=90346117; PubMed=2384148;
CC	Dancer S.J., Garrat R., Saldanha J., Jhoti H., Evans R.;
CC	"The epidermolytic toxins are serine proteases.";
CC	FEBS Lett. 268:129-132(1990).
CC	-I- FUNCTION: HAS SERINE PROTEASE-LIKE PROPERTIES AND BINDS TO THE SKIN PROTEIN PROFILAGGRIN.
CC	-!- DISEASE: EXPOLIATIVE TOXINS CAUSE IMPETIGOUS DISEASES COMMONLY REFERRED AS STAPHYLOCOCAL SCALDED SKIN SYNDROME (SSSS).
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.

CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>).

DR EMBL: M13460; ACC#Z00201.1;
DR PIR: B26680; PRSAEB.
DR PIR: A26050; A26050.
DR HSSP: P09331; IAGJ.
DR

[illegible]

DR	PROSITE; PS00672; V8_SER; 1.
DR	PROSITE; PS00673; V8_SER; 1.
KW	Hydrolase; Serine protease; Toxin; Signal; Plasmid.
FT	SIGNAL 1 31

[illegible]

SEQUENCE 277 AA; 30703 MW; DE1AT/DE127/SC12AT CRO64;
Query Match 8.5%; Score 133; DB 1; Length 277;
Best Local Similarity 28.7%; Pred. No. 0.0097;

BEST LOCAL SIMILARITY 28.7%; PRED NO. 0.0097;
Matches 56; Conservative 26; Mismatches 79; Indels 34; Gaps 9;

Qy 96 TKVTNTRVAPNSIAVIITFGSS-CTCGLIAPNKILTNGHCYNTATRSYSAKGSVPKGM 154
Dh :|:: |::::: :: ||||| :|||| :|::| :

bh 57 THTDMDARNSPGVGTVFKSGSLTATCVLTCKNTIVTNHYH-VAREAAKNPS-----NI 108

DB			57 THITDNARSPNSVGVTVFKSGTGLATGCKNTIVNYH-VAREAAKNPS-----NI	108
	QY	155 NDSTAVNGSANMTFFYPVPGIINT---GASOY----DFAVIKTTDTN-IGNTVGYRSIRQV		206
		:	: : : : : : :	
DB		109 TPDNDANDFAKFNPDPDYCEAFETIKSDYGOCIDIAIKIKDKPKCFAC-DLTD		166

Db 109 IFTPAQRDAEKNEFFPPYCKFEAEIKESPYGGGLDLAIILKLPNKEGSAG--DLLQP 166
:
QY 207 TNL-----TGTTIKISGYPGDKMRSTGKVSEWMSGPVTREDTNLAYITIDTFSGNSG 259

Db	167	ANIPDHIDIOKDKYSLGYPYNSAYSLYQSQIEM-----FNDSQYFGYTEVGN	218
Qy	260	SAMLDNQOQIVGVHN	274

Db 219 SGIFNLKXGELIGHS 233

```

RESULT      7
OMPB_RICJA  OMPB_RICJA
ID          ID          STANDARD;      PRT;  1656 AA.
AC          AC          006653;

```

AC	O06653:
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)

DOI: 10.1001/2001 (REV. 40, last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (5) (Sca5) (rOmpB)
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB003681; BAA20138.1;
CC InterPro: IPR003858; OmpA_rOmpB.
CC Pfam: PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
Query Match 7.9%; Score 123; DB 1; Length 1656;
Best Local Similarity 23.3%; Pred. No. 0.44;
Matches 90; Conservative 52; Mismatches 105; Indels 140; Gaps 25;
QY 4 VKKLIPSLVFG--ALSVPSFAHAASDVL---TSYDMVTSDG----- 42
Nb 8 LKKLISAGLVTAATVAFSAGSAMGAATQONRTNGVATTVDGVGDTVALANVAVA 67
43 -KVISSADFNH--DMKTPS-SFD-----KYDDLSSSTIGK-----VKPLTTLKDFOTKV 88
68 PNAVITANANNINGINLNPAGSFNGFLSNANNLAVTVSEDTTLCGFINNAANNANRFLNLT 127
89 VIGDGRF-----KVTNTR-----VAPYN-----SIAYITGGSSCT- 120
128 ---DAGKTLITGGITNVQSAATHNAQNTIVAKFNGCGAATAMNDLSGLGTFDGAASSTL 184
121 -----GTLIAPNKILNGHCYVNTATRSYSKAGSVYPGMNDSTAVNGSANMTE 168
185 VFDLANPTTOKAPLIDNADNLIVNG-----ANGTLNVT- 217
169 FYVPSGYINTGASQYDFAVIKTDFNIGNTNGYRSIROVTNLT-----GTTIKISGYPG 221
218 ---NGFIQ--VSDKSFATVKA--INIGDGGFMENTNATNANLNQAGGTTINFNGTDG 270
222 DKMRSTGKVSQWEMSGPVTREDTNLAYITDFTSGN--SGSAMLQD---NQIIVGVHNAGY 277
271 -----TGRVLLSLKNGAAT--DFNVT---GSLGSLNKGIIELNTVAINGOLIA--NAGP 317
QY 278 SN-----GTINGGPKATAAFVEFINTAKA 301

Db 318 ANAVICTNNAGRA-AGFVVVVDNGKA 343
RESULT 8
TRY3_AEDAE
ID TRY3_AEDAE STANDARD; PRT; 254 AA.
AC P29786;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin 3Al precursor (EC 3.4.21.4).
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Aedes.
OX NCBI_TaxID=71159;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9742546; PubMed=9087545;
RA Kalhok S., Tabak L.M., Prosser D.E., Brook W., Downer A.E.R.,
RA White B.N.;
RT "Isolation, sequencing and characterization of two cDNA clones coding
RT for trypsin-like enzymes from the midgut of Aedes aegypti";
RL Insect Mol. Biol. 2:71-79(1993).
CC -!- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD
CC MEAL.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: MIDGUT.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X64362; CAA45714.1;
CC PIR: S19890; TRWV3Y.
CC HSSP: P00763; LDPO.
CC MEROPS: S01.112;
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS00240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Digestion;
KW Multigene family.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 27 ACTIVATION PEPTIDE.
FT CHAIN 28 254 TRYPSIN 3Al.
FT ACT_SITE 68 68 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 113 113 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 53 69 BY SIMILARITY.
FT DISULFID 178 194 BY SIMILARITY.
FT DISULFID 205 229 BY SIMILARITY.
FT SITE 203 203 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 254 AA; 26939 MW; 43AAFL3B8DE7D485 CRC64;
Query Match 7.7%; Score 120; DB 1; Length 254;
Best Local Similarity 25.5%; Pred. No. 0.073;
Matches 48; Conservative 31; Mismatches 73; Indels 36; Gaps 10;
QY 115 GGSCTGTLIAPNKLNGHCYVNTATRSYSKAGSVYPGMNDSTAVNGSANM-TEFYVPS 173

```
Db 49 GRHFCGSIISPRWVLTTRAHCTNTDPAAYTIRA-----GSTDRT--NGGIIVKVVSVIPH 102
QY 174 GYINTGASQYDFAVIKTNTDNIGTVGYSIROV-----TNLTGTTIKISGYPGDK--- 223
Db 103 POYNGDYNDFSLLEDESIGFS---RSIEALPDASETVADGAMCTVSGMGDKKNVF 159
QY 224 -----MRSTG--KVSQWMSG-----PVTRDTNLAAYT--IDTFSGNSGSAMLDQNOO 268
Db 160 ENMTLLRAVNPSPYNOACAAALVNVVPVTEQMICAGYAAGKQKDSQSGSGLPV-SGDK 218
QY 269 IVGVHNAQ 276
Db 219 LVGVVSWG 226

LT 9
MYCTU
YF48_MYCTU STANDARD; PRT; 678 AA.
Q10778;
01-OCT-1996 (Rel. 34, Created)
15-JUL-1999 (Rel. 38, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV1548c.
GN RV1548C OR MT1599 OR MTCr48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feilwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*;
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RX Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RX Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RX Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RX Bishai W.;
RT *Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74020; CA98335.1; -
CC EMBL; AE007026; AAK45866.1; ALT_INIT.
CC TIGR; MT1599; -
CC TubercuList; RV1548c; -
CC InterPro; IPR002989; Mycobact_pentapep.
CC InterPro; IPR000030; PPE.
CC Pfam; PF01469; Pentapeptide_2; 13.
CC Pfam; PF00823; PPE; 1.
```

```
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

Query Match 7.6%; Score 119; DB 1; Length 678;
Best Local Similarity 25.7%; Pred. No. 0.29;
Matches 80; Conservative 41; Mismatches 112; Indels 78; Gaps 18;

QY 14 FGALSVPS-----FAHAASDSV--LTSDDYDMV-----TSDGKVVISSADFNDMKTTPSSFDKV 63
Db 379 FGALNSGSGNLGFGNSGNGNIGFNFSGNNIGMNSGNGVGLSVFSGSAERSGFGNS 438
QY 64 DDLSSITGEKVKVPLTTLTKDFOTKVVIGDGRK-----VTNRVAPYNSIAYITFGG 116
Db 439 GELSTGIGNSGQLSTGWFNSATTSTGWFNSGTTNTGWFNSGTTNTGIG--NS-----GG 490
QY 117 SSCTGTLIAPNKLITNGHCYVNTATRSYSAKGSVYPMNDSTAVNGSANMTFFYVPSGYI 176
Db 491 NLVTGSM-----GLFNSGH-----TNTGSFNA-GSM-----NTGDFNSGNNVTGYFNSGNI 535
QY 177 NTGASQYDFAVIKTDTNIG--NTVGYRSIRQVNTLTG-----TTIKISGYPGDKMRSTGKVS 231
Db 536 NTGF-----FNSGDLNTGLFNSYVNPQVNSGWLHTGTNNISGYANAGTFNSG--- 581
QY 232 QWEMSGPVTRDTNLAAYTIDTFSGNSGSAML-DONQOI--VGVNAGYSNG--TINGGP 286
Db 582 -----FDNNARDEHAEPVGTGNSGLANVGNAGIIVGDHLSGFRNSVPTITG-- 629
QY 287 KATAAFVEFIN 297
Db 630 --TANISGFVN 638

RESULT 10
TRYP_SACER
ID TRYP_SACER STANDARD; PRT; 227 AA.
AC P24664;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Trypsin (EC 3.4.21.4) (SET).
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE.
RA Miyamoto K., Matsuo H., Narita K.;
RL (In) Dai 30 kai tanpakushitsu kouzou touronkai kouen youshishuu,
RL pp.77-80 (1979).
RN [2]
RP REVISIONS.
RX MEDLINE=92176171; PubMed=1794983;
RA Nagamine Y., Norioka S., Sakiyama F.;
RL Unpublished results, cited by:
RL Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,
RL Kawata Y., Sakiyama F.;
RL J. Biochem. 110:945-950(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=92176171; PubMed=1794983;
RA Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,
RA Kawata Y., Sakiyama F.;
RT *Crystal structure of Streptomyces erythraeus trypsin at 2.7-A
RT resolution.*;
RL J. Biochem. 110:945-950(1991).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- MISCELLANEOUS: HARDLY AUTOLYZES ITSELF AT ALL AT ITS ACTIVE PH
CC RANGE.
```

```
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC HSP; P20231; IAAO.
DR MEROPS; S01.102; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; TRYp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT ACT_SITE 42 42 CHARGE RELAY SYSTEM.
FT ACT_SITE 88 88 CHARGE RELAY SYSTEM.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM.
FT DISULFID 27 43 CHARGE RELAY SYSTEM.
FT DISULFID 150 164
FT DISULFID 175 199
FT SITE 173 173
SEQUENCE 227 AA; 23308 MW; D5AC5E47B227B418 CRC64;

Query Match 7.6%; Score 118.5; DB 1; Length 227;
Best Local Similarity 25.0%; Pred. No. 0.082;
Matches 54; Conservative 33; Mismatches 82; Indels 47; Gaps 12;

QY 88 VVIGDGRKVTNTRVAPYNSIAVITFGSS-CTGTLIAPNKLITNGHCYVNTATRSYSA 146
DB 1 IVGED-----ANVDHPF-TVALVTPDGOQFCGGLAAPNKKVYTAHCTVG---SOPA 50

QY 147 KGSYPGPN-DSTAVNGSANMTEFFYVPSYINTGASQYDFAVIKTDTNI-----GN 196
DB 51 DINVSGRTVMSSNIGTVSKVNTVWVHPY-QDAAKGFDSVLTLEAPVKEAPIELAKAD 109

QY 197 TVGYSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRDT-NLAY--YT--- 250
DB 110 DAGYAPDTAATILGWNTSGGQADHLQ-----KATVNSDDTKQOAYGEXTPNA 161

QY 251 -----IDTFSGNSGSAMLDQNOOIVGVHNAG 276
DB 162 MVCAGVEGGVDTCQDGGSGPMV-VNNKLIGVTSWG 196

RESULT 11
TRYP_FUSOX STANDARD; PRT; 248 AA.
AC P35049;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
Fusarium precursor (EC 3.4.21.4).
Fusarium oxysporum.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=93324511; PubMed=8332590;
RA Rypniewski W.R., Hastrup S., Betzel C., Dauter M., Dauter Z.,
Papendorf G., Branner S., Wilson K.S.;
RT "The sequence and X-ray structure of the trypsin from Fusarium
oxysporum.";
RL Protein Eng. 6:341-348(1993).
RN [2]
X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).
RA Rypniewski W.R., Dambmann C., von der Osten C., Dauter M.,
Wilson K.S.;
RT "Structure of inhibited trypsin from Fusarium oxysporum at 1.55 A.";
Acta Crystallogr. D 51:73-84(1995).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-xaa, Lys-I-xaa.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
```

```
CC TRYPsin FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; S63827; AAB27568.1;
DR PDB; 1TRY; 01-JAN-96.
DR MEROPS; S01.103; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; TRYp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; zymogen; Signal; 3D-structure.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 25 ACTIVATION PEPTIDE.
FT CHAIN 26 248 TRYPSIN.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM.
FT ACT_SITE 204 204 CHARGE RELAY SYSTEM.
FT DISULFID 50 66
FT DISULFID 174 189
FT DISULFID 200 225
FT SITE 198 198
SEQUENCE 248 AA; 24576 MW; 1A0EBA8C3E70294 CRC64;

Query Match 7.6%; Score 118; DB 1; Length 248;
Best Local Similarity 21.7%; Pred. No. 0.099;
Matches 46; Conservative 42; Mismatches 88; Indels 36; Gaps 9;

QY 115 GGSSTGTLIAPNKLITNGHCYVNTATRSYSAKGSYVPGMNDSTAVNGSANMTEFFYVPSG 174
DB 46 GGPWCGSLLNANTVLTAAHCVSAGYAGSGFOIRA---GSLRSTGGITSLSSVRVHPS 101

QY 175 YINTGASQYDFAVIKTDTNI--GNTVGYRSIRQVTN--LTGTTIKISGYPG-DKMRSTGK 229
DB 102 Y---SGNNDLAILKILSTIPSGNIGYARLAASGSDPVAGSATVAGWGATSEGGSSTP 158

QY 230 VSQWEMSGPVTRDTNLAAYTI-----DTFSGNSGSAMLDQNOOIVGV 272
DB 159 VNLKVTVPVIVSRATRAQVGTSAITNOMFCAGVSSGGKSDCGDGGPIVDSNTLIGA 218

QY 273 HNAGYSNG-----TINGPKATAAFVEFIN-YA 299
DB 219 --VSWGNGCARPNYSGVYASVGALRSFIDTYA 248

RESULT 12
SERI_BOMMO STANDARD; PRT; 389 AA.
ID SERI_BOMMO
AC P07856;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sericin precursor (Silk gum protein).
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Middle silk gland;
RC MEDLINE=83082839; PubMed=6294094;
```


Search completed: June 21, 2002, 18:46:01
Job time: 220 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:38:51 ; Search time 12.97 Seconds
(without alignments)
568.738 Million cell updates/sec

Title: US-09-652-743A-12
Perfect score: 1559
Sequence: 1 MKVKKLIPSLVFGALVSP.....NGGPKATAAFVEFINYAKAQ 302

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
arched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	405.5	26.0	222	1 US-08-090-048-1	Sequence 1, Appl
2	405.5	26.0	222	2 US-08-292-550-1	Sequence 1, Appl
3	405.5	26.0	222	2 US-07-927-661A-1	Sequence 1, Appl
4	184	11.8	344	1 US-08-657-192-3	Sequence 3, Appl
5	184	11.8	344	3 US-08-523-373-5	Sequence 5, Appl
6	184	11.8	392	3 US-08-523-373-6	Sequence 6, Appl
7	184	11.8	532	1 US-08-657-192-9	Sequence 9, Appl
8	184	11.8	532	3 US-08-523-373-7	Sequence 7, Appl
9	184	11.8	537	1 US-08-657-192-15	Sequence 15, Appl
10	183	11.7	213	3 US-08-523-373-22	Sequence 22, Appl
11	183	11.7	214	3 US-08-523-373-23	Sequence 23, Appl
12	183	11.7	215	3 US-08-523-373-24	Sequence 24, Appl
13	182	10.4	241	4 US-08-896-933-32	Sequence 32, Appl
14	182	10.4	241	4 US-09-314-235-32	Sequence 32, Appl
15	134	8.6	246	4 US-08-896-933-33	Sequence 33, Appl
16	134	8.6	246	4 US-09-314-235-33	Sequence 33, Appl
17	129.5	8.3	383	4 US-09-072-384-18	Sequence 18, Appl
18	129.5	8.3	392	4 US-09-072-384-15	Sequence 15, Appl
19	121.5	7.8	256	4 US-09-032-215-27	Sequence 27, Appl
20	121.5	7.8	392	4 US-09-072-384-2	Sequence 2, Appl
21	120	7.7	255	3 US-08-906-613-83	Sequence 83, Appl
22	120	7.7	255	3 US-08-906-616-83	Sequence 83, Appl
23	120	7.7	255	4 US-08-817-795-83	Sequence 83, Appl
24	120	7.7	255	4 US-08-639-075A-83	Sequence 83, Appl
25	120	7.7	255	4 US-09-012-431-83	Sequence 83, Appl
26	120	7.7	255	4 US-09-012-692-83	Sequence 83, Appl
27	120	7.7	255	4 US-08-906-613-83	Sequence 83, Appl

Query Match 26.0%; Score 405.5; DB 1; Length 222;
Best Local Similarity 39.3%; Pred. No. 3e-29;

28	120	7.7	255	5 PCT-US95-14442A-83	Sequence 83, Appl
29	118	7.6	224	1 US-08-553-516-2	Sequence 2, Appl
30	118	7.6	248	1 US-08-238-130-2	Sequence 2, Appl
31	118	7.6	248	2 US-08-921-426-4	Sequence 4, Appl
32	118	7.6	248	3 US-08-816-915-4	Sequence 4, Appl
33	118	7.6	248	5 PCT-US95-07743-4	Sequence 4, Appl
34	106	6.8	245	3 US-08-906-769-121	Sequence 121, App
35	106	6.8	245	3 US-08-906-616-121	Sequence 121, App
36	106	6.8	245	4 US-08-639-075A-121	Sequence 121, App
37	106	6.8	245	4 US-09-012-431-121	Sequence 121, App
38	106	6.8	245	4 US-09-012-692-121	Sequence 121, App
39	106	6.8	245	4 US-08-906-613-121	Sequence 121, App
40	105	6.7	653	4 US-08-849-602C-27	Sequence 27, Appl
41	105	6.7	653	6 5248599-2	Patent No. 5248599
42	104.5	6.7	255	1 US-08-650-129-5	Sequence 5, Appl
43	104.5	6.7	255	3 US-08-984-417-5	Sequence 5, Appl
44	102	6.5	1536	1 US-08-038-682-2	Sequence 2, Appl
45	102	6.5	1536	1 US-08-302-832-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-090-048-1
: Sequence 1, Application US/08090048
: Patent No. 5523237
: GENERAL INFORMATION:
: APPLICANT: Budtz, Peter
: TITLE OF INVENTION: PROTEIN PREPARATIONS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5523237o No. 5523237disk of No. 5523237th America, Inc.
: STREET: 405 Lexington Avenue, 62nd Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/090,048
: FILING DATE: 16-JUL-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 199/91
: FILING DATE: 06-FEB-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/DK92/00036
: FILING DATE: 06-FEB-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Agis, Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 3396.214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-867-0298
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 222 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-090-048-1

```

Matches 83; Conservative 36; Mismatches 77; Indels 15; Gaps 5
QY 89 VIGDGRKTVNTRVAPYNSIAYITFGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKG 148
   ||| | ||| ||| || | : | : ||| | : | : ||| | : | : ||| | : | :
Db 2 VIGSDRTRVNTTAYPYRAIVHISIGSCTGWMIGPKTVATAGHCIIYDTSSGSFAGTA 61
QY 149 SVTPGMNDSTAVNGSANMTEFYVPSGINTGASOYDFAVIKTDNIGHTVGYRSIRQVT- 207
   : || | : || | : || | : || | : || | : || | : || | : || | : || |
Db 62 TVSPGRNGTSPYGVSKSTRYFIPSGW-RSGNTNDYGAIEIELSEPIGNTVGYFGYSYTS 120
QY 208 NLTGCTTIKISGCDKMRSTCKVSWQWMSGPVTREDTNLAYTTIDTSGNSGSAMLQDQ 267
   : || | : ||| ||| || | : || | : || | : || | : || | : || | : || |
Db 121 SLVGTGTTVTSIGYPGDKTAGT----QWHRSGGAISETSEYKLOYAMDITVGGQSGPVFEQSS 176
QY 268 Q-----TVGVHNAG-YSNGTINGGPKAT 289
   : || | : || | : || | : || | : || | : || | : || | : || |
Db 177 SRTNCSGPCSLAVHTNGVYGGSSYNRGTRIT 207
-----LT 2
'-292-550-1
ence 1, Application US/08292550
cent No. 5863573
; GENERAL INFORMATION:
; APPLICANT: Dambmann, Claus
; APPLICANT: Budtz, Peter
; APPLICANT: Mortensen, Steen B.
; TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58635730 No. 5863573disk of No. 5863573th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #F.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/292,550
; FILING DATE: 18 AUG 1994
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/927,624
; FILING DATE: 31 AUG 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/0068
; FILING DATE: 8 MAR 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-292-550-1
Query Match 26.0%; Score 405.5; DB 2; Length 222;
Best Local Similarity 39.3%; Pred. No. 3e-29;
Matches 83; Conservative 36; Mismatches 77; Indels 15; Gaps 5
QY 89 VIGDGRKTVNTRVAPYNSIAYITFGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKG 148
   ||| | ||| ||| || | : | : ||| | : | : ||| | : | : ||| | : | :

```

```

Db      2  VIGSDRTRVNTATPYRAIVHISSIGSCTGWMIGPKTVATAGHCIIYDTSSGSGFAGTA 61
Qy      149  SVTPGMDNSTAVANGSANMTFEFYVPSGYINTGASQYDFAVIKTDNTNIGNTVGYRSIRQVT- 207
      :| | | | :| | | | | :| | | | :| | | | :| | | | |
Db      62  TVSPGRNGTSYPPYGSVKSTRYFIPSCW-RSGNTNIYDGAIELSEPIGNTVGYFGSYTTS 120
Qy      208  NLGTGTTIKISYPCDKMRSTGKVSQWEMGGPVTREDTNLAYTYTIDTFSGNSGSAMLDDNQ 267
      :| | | | | | | | | | | | | | :| | | | :| | | | :| | | | :| | | |
Db      121  SLVGTGTVTSYGPCDKTAGT- ---QWHSGPATSETYKLYQVAMDYTGQSGSPVFEQSS 176
Qy      268  Q-----IVGVHNAG-YSNGTINGGPKAT 289
      :| | | | | | | | | | | | | |
Db      177  SRTNCSPCSLAVHTNGVYGGSSYNRGTRIT 207

RESULT      3
; Sequence 1, Application US/07927661A
; Patent No. 5866357
; GENERAL INFORMATION:
; APPLICANT: Dambmann, Claus
; APPLICANT: Mortensen, Steen
; APPLICANT: Budtz, Peter
; APPLICANT: Eriksen, Svend
; TITLE OF INVENTION: Protein Hydrolysates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866357o No. 5866357disk of No. 5866357th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/927,661A
; FILING DATE: 31-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3396.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5866357e
; US-07-927-661A-1

```

	Query Match	26.0%	Score 405.5	DB 2	Length 222
	Best Local Similarity	39.3%	Pred. No. 3e-29		
	Matches 83	Conservative 36	Mismatches 77	Indels 15	Gaps 5
Qy	89	VIGDGRKVTNTRVAPVNSIAYITFGSSCTGTLIAPNKILTNCHGVYNTATRSYAKG	148		
Db	2	VIGSDRTRVNTTAYPYRAIVHVISSISGSSCTGMWIGPKTVATAGHCIIYDTSGSFAGTA	61		
Qy	149	SVYFGMDSSTAVGNSANNTTEYVPSGYNTCASQYDFAVIKTDINIGNTCYRSIQVT	207		
Db	62	TVSPGRNGTSYPYGVSKSTRFIPSGWRSNGNTIDYGAIELSEPIGNTVGFVGYSTTS	120		
Oy	208	NLTGTTIKISGYPGDKMRSTCKVSWENSGSPVTRDTNLAYITTYITDTSFGNSGSAMLDQO	267		

Db 121 SLVGTITVIGPGDKTACT-----QWHSQPIAISTYKLYQYAMDTYGGSGSPVFEQSS 176
QY 268 Q-----IVGVHAG-YSNCTINGGPKAT 289
Db 177 SRTNCSGPCSLAVHTNGVYGGSSYNRGTRIT 207

RESULT 4
US-08-657-192-3
; Sequence 3, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657.192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-657-192-3

Query Match 11.8%; Score 184; DB 1; Length 344;
Best Local Similarity 25.3%; Pred. No. 6.5e-09;
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;
QY 87 KVIIGDGRKVTNTRVAPYNSIAYITF-----GSSCTGTLIAPNKILTNHCHVYNTATR 142
Db 124 EVILPNDRHQITDTTNGHYAPVTYIOVEAPTGTFTIASGVVGVGKDTLLTNKHVVDTATGD 183
QY 143 SYSAGSVYP-GMNDSTAVNG---SANTTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196
Db 184 PHALK--AFPSAINQDNPNGGFTAENITKY-----SGEGLAIVKFSPEQNKH 231
QY 197 -----TVGYSRSTQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTTRETNLAY 248
Db 232 IGEVVKPATMSNNAETQVQ-----NITVTGYPGDKPVAT---MMESKGIITLKGEMQ 283
QY 249 YTIDTFSGNSGAMLDQNOQIIVGVHNGYSNGTINGGPKATAAFV 293
Db 284 YDLSTGTGNSGSPVFNKNEKNEVIGIHWGVPN-EFNG-----AVFI 322

RESULT 6
US-08-523-373-6

RESULT 5
US-08-523-373-5
; Sequence 5, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523.373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-5

Query Match 11.8%; Score 184; DB 3; Length 344;
Best Local Similarity 25.3%; Pred. No. 6.5e-09;
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;
QY 87 KVIIGDGRKVTNTRVAPYNSIAYITF-----GSSCTGTLIAPNKILTNHCHVYNTATR 142
Db 124 EVILPNDRHQITDTTNGHYAPVTYIOVEAPTGTFTIASGVVGVGKDTLLTNKHVVDTATGD 183
QY 143 SYSAGSVYP-GMNDSTAVNG---SANTTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196
Db 184 PHALK--AFPSAINQDNPNGGFTAENITKY-----SGEGLAIVKFSPEQNKH 231
QY 197 -----TVGYSRSTQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTTRETNLAY 248
Db 232 IGEVVKPATMSNNAETQVQ-----NITVTGYPGDKPVAT---MMESKGIITLKGEMQ 283
QY 249 YTIDTFSGNSGAMLDQNOQIIVGVHNGYSNGTINGGPKATAAFV 293
Db 284 YDLSTGTGNSGSPVFNKNEKNEVIGIHWGVPN-EFNG-----AVFI 322

```

; APPLICANT: OHSUYE, Kazuhiro
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-657-192-9

Query Match 11.8%; Score 184; DB 1; Length
Best Local Similarity 25.3%; Pred. No. 1.2e-08;
Matches 57; Conservative 41; Mismatches 83; Index

QY 87 KVIIGDGRKVTNTRVAPYNSIAYITF----GGSSCTGTLIAPNKLITL
DB 124 EVILPNDRHQIDTTNGHYAPTYIQVEAPTGTFTASGVVGVCKDLITL
QY 143 SYSAGSVYP-GMNDSTAVNG----SANMTFFYPVPSYINTGASQYDFAVI
DB 184 PHALK--AFPSAINQDNPYNGGFTAENITKY-----SGEGDLAIV
QY 197 -----TVGYRSIROVTNLTGTTIKISYCPGDKMRSTCKVSWQMSG
DB 232 IGEVVKPATMNSNAETQVNO-----NITVTGYPGDKPVAT----MWESKGF
QY 249 YTIDTFSGSGSAMLQDNOQIVGHVHAGYSNGTINGSPKATAAFV 293
DB 284 YDLSTTCGNSGSPVFNKNEVIGHGCVPN-EPNG-----AVFI 322

RESULT 8
US-08-523-373-7
; Sequence 7, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street

```

CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-7

Query Match 11.8%; Score 184; DB 3; Length 532;
Best Local Similarity 25.3%; Pred. No. 1.2e-08;
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;
QY 87 KVVIGDGRKVTNTRVAPYNSIAYITF----GGSSCTCTLIAPNPKILTNHGCYVNTAPR 142
DB 124 EVILPNDRHQITDTTNHGYAPVTYIQVEAPTGTFTASGVVYVVKDTLLTNKKHVDATHGD 183
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDITNIG-- 196
DB 184 PHALK--APPSAINQDNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231
QY 197 -----TVGYSRISQVNTNLGTTIKISGYFGDKMRSTGKVSOWEMSGPVTREDTNLAY 248
DB 232 ICEVVKPATMSNNAETOVNQ----NITVTGYFGDKPVPAT-----AVFI 322
QY 249 YTIIDTFSGNSAMLDONQOIVGVHNAGYSNGTINGGPKATAAFV 293
DB 284 YDLSTTGGNSGSPVFNKNEVIGIHWGGVYPN-EFNG-----AVFI 322

RESULT 9
US-08-657-192-15
Sequence 15, Application US/08657192
Patent No. 5747321
GENERAL INFORMATION:
APPLICANT: YABUTA, Masayuki
APPLICANT: OHSUYE, Kazuhiro
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
TITLE OF INVENTION: PROTEASES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States

ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/657,192
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-170086
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-15

Query Match 11.8%; Score 184; DB 1; Length 537;
Best Local Similarity 25.3%; Pred. No. 1.2e-08;
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;
QY 87 KVVIGDGRKVTNTRVAPYNSIAYITF----GGSSCTCTLIAPNPKILTNHGCYVNTAPR 142
DB 124 EVILPNDRHQITDTTNHGYAPVTYIQVEAPTGTFTASGVVYVVKDTLLTNKKHVDATHGD 183
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDITNIG-- 196
DB 184 PHALK--APPSAINQDNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231
QY 197 -----TVGYSRISQVNTNLGTTIKISGYFGDKMRSTGKVSOWEMSGPVTREDTNLAY 248
DB 232 ICEVVKPATMSNNAETOVNQ----NITVTGYFGDKPVPAT-----AVFI 322
QY 249 YTIIDTFSGNSAMLDONQOIVGVHNAGYSNGTINGGPKATAAFV 293
DB 284 YDLSTTGGNSGSPVFNKNEVIGIHWGGVYPN-EFNG-----AVFI 322

RESULT 10
US-08-523-373-22
Sequence 22, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhiro
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-22

Query Match 11.7%; Score 183; DB 3; Length 213;
Best Local Similarity 25.4%; Pred. No. 4.1e-09;
Matches 57; Conservative 83; Mismatches 83; Indels 44; Gaps 10;

QY 88 VVIGDGRKTVNTRVAPYNSIAIYTF-----GGSSCTGTLLIAPNKLITNGHCVYNTATRS 143
DB 1 VILPNDRHQITDTNGHYAPVTYIQVEAPTGTFTIASGVVVGKDTLLTNKHVVVDATHGDP 60

QY 144 YSAKGSVYP-GMNDSTAVNG---SANTFEYVPSGYINTGASOYDFAVIKTDNIGN--- 196
DB 61 HALK--APPSAINODNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKHI 108

QY 197 -----TVGYRSIRQVNTLTGTIKISYPGDKMRSTGKVSOWEMSGPVTREDTNLAYY 249
DB 109 GEVVKPATMSNNAETQVNO---NITVTGPGDKPVAT-----MWESKGIYTLKGEAMQY 160

QY 250 TIDTFSGNSGSMLDQNOQIVGVHNAGYSNGTINGGPKATAAFV 293
DB 161 DLSTTGGSGSPVNEKNEVIGIHWGVNP-EFNG-----AVFI 198

RESULT 11
US-08-523-373-23
Sequence 23, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhiro
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 214 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-23

Query Match 11.7%; Score 183; DB 3; Length 214;
Best Local Similarity 25.4%; Pred. No. 4.1e-09;
Matches 57; Conservative 40; Mismatches 83; Indels 44; Gaps 10;

QY 88 VVIGDGRKTVNTRVAPYNSIAIYTF-----GGSSCTGTLLIAPNKLITNGHCVYNTATRS 143
DB 1 VILPNDRHQITDTNGHYAPVTYIQVEAPTGTFTIASGVVVGKDTLLTNKHVVVDATHGDP 60

QY 144 YSAKGSVYP-GMNDSTAVNG---SANTFEYVPSGYINTGASOYDFAVIKTDNIGN--- 196
DB 61 HALK--APPSAINODNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKHI 108

QY 197 -----TVGYRSIRQVNTLTGTIKISYPGDKMRSTGKVSOWEMSGPVTREDTNLAYY 249
DB 109 GEVVKPATMSNNAETQVNO---NITVTGPGDKPVAT-----MWESKGIYTLKGEAMQY 160

QY 250 TIDTFSGNSGSMLDQNOQIVGVHNAGYSNGTINGGPKATAAFV 293
DB 161 DLSTTGGSGSPVNEKNEVIGIHWGVNP-EFNG-----AVFI 198

RESULT 12
US-08-523-373-24
Sequence 24, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhiro
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028

;; FILING DATE: 07-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Neuth, Donna M.
;; REGISTRATION NUMBER: 36,607
;; REFERENCE/DOCKET NUMBER: 001560-251
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-836-6620
;; TELEFAX: 703-836-2021
;; INFORMATION FOR SEQ ID NO: 24:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 215 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; 78-523-373-24

Query Match 11.78; Score 183; DB 3; Length 215;
Best Local Similarity 23.48; Pred. No. 4.le-08;
Matches 57; Conservative 40; Mismatches 83; Indels 44; Gaps 10;

Qy 88 VVIGDGRTKVNTVAIPYNSIATITF---GGSSCTGTLIAPNKLITNGHCYVNTATRS 143
Db 1 VILPNDRHQITDTHGHAPYTIQVEAPTGTFTIASGVVYVCKDTLLTNKHHVDATHGDP 60

Qy 144 YSAKGSVYP-GMNDSTAVNG---SAMTFEYVPSGYINTGASQYDFAVIKTDTNIGN--- 196
Db 61 HALK--AFPSAINODNYPNGGTAENITKY-----SGEGDLAIVKFSPEQNKH 108

Qy 197 -----TVGYSIRIQVNTLGTITKISGYPCDKMRSTGKVSOWEMSGPVTREDTNLAY 249
Db 109 GEVVKPATMSNAETQVQV---NITVTGYCDKPVAT---MMESKGIITLKGAEAMQY 160

Qy 250 TIDTFSGNSGAMLQOQIVGVHNGYSNGTINGSGPKATAAFV 293
Db 161 DLSTTGGNSGSPVFNKNEKVEIGHWGGVNP-EFNG-----AVFI 198

RESULT 13
US-08-896-933-32
; Sequence 32, Application US/08896933
; Patent No. 6221351
; GENERAL INFORMATION:
; APPLICANT: Terman, David S.
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS
; FILE REFERENCE: 09629/005002
; CURRENT APPLICATION NUMBER: US/08/896,933
; CURRENT FILING DATE: 1997-07-18
; EARLIER APPLICATION NUMBER: 08/252,978
; EARLIER FILING DATE: 1994-06-02
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-08-896-933-32

Query Match 10.48; Score 162; DB 4; Length 241;
Best Local Similarity 30.08; Pred. No. 3.9e-07;
Matches 62; Conservative 26; Mismatches 93; Indels 26; Gaps 9;

Qy 105 PYNSTAYITFCGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKGSVYPCMDNDSTAVNGSA 164
Db 42 PYNTIGNVFKGTSATGLICKNTVLTNRHIA--KFANGDPSKVSFRFSINTDDNGNTET 99

Qy 165 NMTEFYVPSGYINTGASQYDFAVI--KTDN---IGNTVGYSIRIQVNTL-TGTTIKISG 218
Db 100 PYGEYEVKEILQEPFGAGVDLALRLKPDQNGVSLGDKISPAKIGTSDNLKDGKLELIG 159

Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSGNSGAMLQOQIVGVHNGY 277
Db 160 YPFDH-----KVNQMRSEIELTTLRGLRYGFTVPCNSGSGIFNSGELVGIHSSKV 213

Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300
Db 214 SHLDREHQINYG-----VGIGNYVK 233

Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSGNSGAMLQOQIVGVHNGY 277
Db 160 YPFDH-----KVNQMRSEIELTTLRGLRYGFTVPCNSGSGIFNSGELVGIHSSKV 213

Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300
Db 214 SHLDREHQINYG-----VGIGNYVK 233

RESULT 14
US-09-314-235-32
; Sequence 32, Application US/09314235
; Patent No. 6338845
; GENERAL INFORMATION:
; APPLICANT: Terman, David S.
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS
; FILE REFERENCE: 09629/005004
; CURRENT APPLICATION NUMBER: US/09/314,235
; CURRENT FILING DATE: 1999-05-18
; EARLIER APPLICATION NUMBER: 08/896,933
; EARLIER FILING DATE: 1997-07-18
; EARLIER APPLICATION NUMBER: 08/252,978
; EARLIER FILING DATE: 1994-06-02
; EARLIER APPLICATION NUMBER: 07/891,718
; EARLIER FILING DATE: 1992-06-01
; EARLIER APPLICATION NUMBER: US91/00342
; EARLIER FILING DATE: 1991-01-17
; EARLIER APPLICATION NUMBER: 07/466,577
; EARLIER FILING DATE: 1990-01-17
; EARLIER APPLICATION NUMBER: 07/416,530
; EARLIER FILING DATE: 1989-10-03
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-314-235-32

Query Match 10.48; Score 162; DB 4; Length 241;
Best Local Similarity 30.08; Pred. No. 3.9e-07;
Matches 62; Conservative 26; Mismatches 93; Indels 26; Gaps 9;

Qy 105 PYNSTAYITFCGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKGSVYPCMDNDSTAVNGSA 164
Db 42 PYNTIGNVFKGTSATGLICKNTVLTNRHIA--KFANGDPSKVSFRFSINTDDNGNTET 99

Qy 165 NMTEFYVPSGYINTGASQYDFAVI--KTDN---IGNTVGYSIRIQVNTL-TGTTIKISG 218
Db 100 PYGEYEVKEILQEPFGAGVDLALRLKPDQNGVSLGDKISPAKIGTSDNLKDGKLELIG 159

Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSGNSGAMLQOQIVGVHNGY 277
Db 160 YPFDH-----KVNQMRSEIELTTLRGLRYGFTVPCNSGSGIFNSGELVGIHSSKV 213

Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300
Db 214 SHLDREHQINYG-----VGIGNYVK 233

RESULT 15
US-08-896-933-33
; Sequence 33, Application US/08896933
; Patent No. 6221351
; GENERAL INFORMATION:
; APPLICANT: Terman, David S.
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS
; FILE REFERENCE: 09629/005002
; CURRENT APPLICATION NUMBER: US/08/896,933
; CURRENT FILING DATE: 1997-07-18

EARLIER APPLICATION NUMBER: 08/252,978
EARLIER FILING DATE: 1994-06-02
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33
LENGTH: 246
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-08-896-933-33

Query Match 8.6%; Score 134; DB 4; Length 246;
Best Local Similarity 28.4%; Pred. No. 0.00014;
Matches 58; Conservative 27; Mismatches 83; Indels 36; Gaps 10;
QY 96 TKVTNTRVAPYNSIAVITFGSS-CTGTLIAPNKILTNHGVYNTATRSYSAKGSVYPGM 154
Db 26 THITDNARSPYNSVGVTVKGSTLATGVLIGKNTIVTYH-VAREAAKNPS-----NI 77
155 NDSTAVNGSANMTEFYVPSGYINT---GASQY----DFAVIKTDTN-IGNTVGYRSIROV 206
78 IFTPAQNRAEKNEFPTYGKFEAEIKESPYGGGLDLAIILKLPNEKESAG--DLIQP 135
QY 207 TNL-----TGTIKISGPGDKMRSTGKVSOWENSGPVTREDTNLAYTIDTFSGNSG 259
Db 136 ANIPRHDIOKDKYSLGYPYNSAYSLSYQSIEM-----FNDSQYFGYTEVGN 187
QY 260 SAMLDQNOOIVGVHNGYNGTIN 283
Db 188 SGIFNLKGLIGIHSG--KGGQMN 209

Search completed: June 21, 2002, 18:42:16
Job time: 205 sec